



The ReadFree tool for the identification of poor readers: a validation study based on a machine learning approach in monolingual and minority-language children

Desiré Carioti^{1,2} · Natale Adolfo Stucchi² · Carlo Toneatto² · Marta Franca Masia¹ · Milena Del Monte^{1,3} · Silvia Stefanelli^{1,4} · Simona Travellini^{1,3} · Antonella Marcelli³ · Marco Tettamanti² · Mirta Vernice¹ · Maria Teresa Guasti² · Manuela Berlinger^{1,3,5}

Received: 9 May 2023 / Accepted: 26 June 2023 / Published online: 7 August 2023
© The Author(s) 2023

Abstract

In this study, we validated the “*ReadFree tool*”, a computerised battery of 12 visual and auditory tasks developed to identify poor readers also in minority-language children (MLC). We tested the task-specific discriminant power on 142 Italian-monolingual participants (8–13 years old) divided into monolingual poor readers ($N = 37$) and good readers ($N = 105$) according to standardised Italian reading tests. The performances at the discriminant tasks of the “*ReadFree tool*” were entered into a classification and regression tree (CART) model to identify monolingual poor and good readers. The set of classification rules extracted from the CART model were applied to the MLC’s performance and the ensuing classification was compared to the one based on standardised Italian reading tests. According to the CART model, auditory go-no/go (regular), RAN and Entrainment_{100bpm} were the most discriminant tasks. When compared with the clinical classification, the CART model accuracy was 86% for the monolinguals and 76% for the MLC. Executive functions and timing skills turned out to have a relevant role in reading. Results of the CART model on MLC support the idea that ad hoc standardised tasks that go beyond reading are needed.

Keywords Bilingualism · Classification and regression tree (CART) model · Developmental dyslexia · Executive functions · Timing skills

✉ Desiré Carioti
desire.carioti@unimib.it

Manuela Berlinger
berlinger@uniurb.it

¹ DISTUM, Department of Humanities, University of Urbino Carlo Bo, Urbino, Italy

² Psychology Department, University of Milano-Bicocca, Milan, Italy

³ Center of Developmental Neuropsychology, AST Pesaro-Urbino, Distretto di Pesaro, Pesaro, Italy

⁴ Department of Human Sciences, University of the Republic of San Marino, San Marino, Republic of San Marino

⁵ NeuroMi, Milan Center for Neuroscience, Milan, Italy

Abbreviations

MLC	Minority-language children
RAN	Rapid automatized naming
CART	Classification and regression tree
GR	Good readers
PR	Poor readers
PCA	Principal component analysis

Introduction

The need for available tests and tools for a fast and accurate screening of reading deficits comes from the high number of children notified by schools and teachers as being at risk of developmental dyslexia and, more in general, of learning disorders. Not all of them receive a fast diagnosis, and once neuropsychologically assessed, not all of them manifest reading disorders. These critical facts add to specific concerns on the neuropsychological assessment of multilingual students, for which there is still a lack of ad hoc created tests and of clinical consensus and criteria, even if literacy difficulties have been documented for first- and second-generation immigrants (Arikan et al., 2017; Bonifacci & Tobia, 2016; Rangvid, 2007; Schnepf, 2004 in Italy by Azzolini et al., 2012; Murineddu et al., 2006). This issue has been recently highlighted in the new “Italian guidelines for the identification of Specific Learning Disorder”, Recommendation 7.3¹. This recommendation (p. 72) reports that “for the identification of Specific Learning Disorders (i.e., Dyslexia and Dysorthography) in a bilingual population, it is recommended to use tests standardised on a bilingual sample”.

To address this issue, we envisaged a computerised screening tool, namely the *Read-Free tool*, capable of detecting behavioural cognitive markers of reading difficulties in both monolinguals and minority-language children (MLC). Our battery minimises the involvement of language processing to obtain cognitive measures free from potential biases associated with use and exposure to variable languages, such as in the case of MLC. Our final goal is to optimise the assessment requests that specialised neuropsychological centres receive, thus reducing the level of burden sustained by National Health Systems.

Markers of developmental dyslexia

Developmental dyslexia is a neurodevelopmental disorder “characterised by problems with accurate or fluent written word recognition, poor decoding, and poor spelling abilities” (DSM-5; American Psychiatric Association, 2013, p. 67). DSM-5 (American Psychiatric Association, 2013) describes the specific learning impairment of readers with dyslexia as an “unexpected” inability to acquire literacy despite preserved reasoning skills. Indeed, to be classified as readers with dyslexia, individuals must show reading skills “below age expectations” despite adequate instruction and educational opportunities, and despite the presence of intact non-verbal reasoning skills, intelligence, and sensory abilities (Peterson & Pennington, 2015). Causes of dyslexia were investigated by several

¹ Linee Guida sulla Gestione dei Disturbi Specifici dell’Apprendimento, June 2021, published by the Italian National Guidelines System, Rome 20 January 2022. https://www.scuolainforma.it/wp-content/uploads/2022/01/LG-389-AIPO_DSA.pdf.

theoretical frameworks (see Carioti et al., 2021; Peterson & Pennington, 2012; Stein, 2018 for reviews) and, accordingly, many empirical findings were provided about dyslexia in different cognitive and perceptual domains. For this reason, dyslexia is described by recent accounts as a multiple deficit disorder, in which different patterns of underlying cognitive deficits can characterise the reading impairment (McGrath et al., 2020; Pennington, 2006; Pennington et al., 2012; Ring & Black, 2018). This view that has been supported by comparative studies (Danelli et al., 2017; Ramus et al., 2003; Reid et al., 2007), although the more replicated findings seem to be those concerning the phonological deficit (Bradley & Bryant, 1978; Elbro & Jensen, 2005; Joanisse et al., 2000; Ramus & Szenkovits, 2008).

Indeed, as widely observed in the literature, beyond the obvious differences in reading tasks, readers with dyslexia may show deficits in cognitive dimensions such as phonological awareness, rapid automatized naming (RAN), verbal working memory (see Carioti et al., 2021 for a review), and also non-verbal abilities related to auditory perception (Goswami et al., 2011; Huss et al., 2011; Tallal, 1980; Tallal & Gaab, 2006; Thomson et al., 2006; Thomson & Goswami, 2008) and visuo-attentional skills (Facoetti et al., 2019; Facoetti & Molteni, 2001; Facoetti, Paganoni, & Lorusso, 2000a; Franceschini et al., 2012). Interestingly, some of these cognitive markers of the disorder (phonological awareness, RAN, verbal working memory) are universal and were, thus, found to be impaired in dyslexia across ages and orthographies (see Carioti et al., 2021 for a review).

Above these most evaluated cognitive skills, several non-verbal paradigms were put forward by experimental psychology to verify causal theories of dyslexia such as the rapid auditory processing theory (Tallal, 1980), the perceptual anchoring theory (Banai & Ahissar, 2010), the temporal sampling theory (Goswami et al., 2011; Huss et al., 2011), and the magnocellular theory (Stein & Walsh, 1997). These non-verbal or language-independent paradigms identified some peculiar behavioural alterations in readers with dyslexia also for perceptual skills as, for example, discrimination of tones/phonemes/syllables (Baldeweg et al., 1999; Fostick et al., 2012; Richardson et al., 2004), rapid auditory sequencing (Georgiou et al., 2010; Hari & Kiesilä, 1996; Laasonen et al., 2001), timing skills (Farmer & Klein, 1995; Flaunacco et al., 2014; Gaab et al., 2007; Tallal & Gaab, 2006; Thomson et al., 2006; Thomson & Goswami, 2008; see Hämäläinen et al., 2013 for a review), motion perception (Hari & Renvall, 2001; Mascheretti et al., 2018; see Benassi et al., 2010 for a review), visuo-attentional skills (Facoetti, Paganoni, Turatto, et al., 2000b), and executive functions tested by means of paradigms such as the Stroop test, the go/no-go task, the Wisconsin card sorting test, and many others (see Booth et al., 2010 and Lonergan et al., 2019 for reviews).

Searching for language-independent tasks: rhythm as a matter of phonology

Except for phonological processing and verbal working memory, some of the cognitive skills mentioned above and considered reliable markers of dyslexia do not necessarily involve language processing and can be tested through language-independent tasks. This is very relevant for the issue addressed by the present work, since the identification of dyslexia-related non-verbal deficits offers the opportunity to find cognitive markers of reading deficits also in a population for whom the “linguistic bias” is an obstacle to diagnosis, as in the case of minority-language students.

A growing body of recent works investigated the relationship between rhythmic skills, language, and the reading disorder (Boll-Avetisyan et al., 2020; Flaunacco et al., 2014; Pagliarini et al., 2020; Thomson et al., 2006; Thomson & Goswami, 2008). Readers with

dyslexia seem to show impairments in rhythmic tasks as maintaining a regular tapping both in childhood (Leong & Goswami, 2014a; Thomson & Goswami, 2008) and adulthood (Leong & Goswami, 2014b; Thomson et al., 2006). These findings also promoted rhythmic and musical intervention for dyslexia (Bonacina et al., 2015; Flaugnacco et al., 2015; Overy, 2003; Thomson et al., 2013), often in a computerised form (Cancer et al., 2020). These language-independent auditory tasks may represent an alternative method to assess phonological difficulties that characterize dyslexia. This idea has been supported by studies that explored the relationship between phonological processing and auditory perception (see the rapid auditory processing theory, Gaab et al., 2007; Tallal, 1980; Tallal & Gaab, 2006), detection of stressed metrical elements in language (see the temporal sampling hypothesis, Goswami et al., 2011; Huss et al., 2011), and rhythmic production conceived as a regular planned movement (see the cerebellar theory, Nicolson et al., 2001a, 2001b; Nicolson & Fawcett, 1990, 2011; Nicolson et al., 1999). In other words, the adoption of non-verbal and, thus, language-independent tasks would allow clinicians to assess the children while overcoming the linguistic and orthographic gap that often prevents an unbiased evaluation of bilingual students (Everatt et al., 2000) and MLC.

Searching for language-independent tasks: executive functions, RAN, and reading skills

In the same line of reasoning, several studies suggested that reading disorders may be associated with poor executive functions abilities (Barbosa et al., 2019; Brosnan et al., 2002; Doyle et al., 2018; Moura et al., 2014; Reiter et al., 2005; Smith-Spark et al., 2016; Varvara et al., 2014). As mentioned above (paragraph 1.1), several language-independent executive functions tasks were tested on readers with dyslexia (see table 2 in Booth et al., 2010, p. 152); some of them, like the well-known go/no-go task (Donders, 1969, in Gomez et al., 2007), can be easily adapted for clinical and screening tests.

The RAN itself, another reliable marker of dyslexia (Denckla & Rudel, 1976; Wolf & Bowers, 1999; see Araújo & Faisca, 2019 for a review), can be easily transposed in a computerised version that, if realised with a limited number of non-alphanumeric stimuli, would considerably reduce language processing, being suitable also to assess MLC (Carioti et al., 2022).

In line with these considerations, we made an effort to develop a computerised screening tool, i.e., the *ReadFree tool*, capable of identifying both monolingual and MLC at risk of reading disorders without using any reading or linguistic tasks, and also taking into account the auditory-visual dichotomy that seems to characterise different dyslexia profiles (Castles & Coltheart, 1993).

The multiple deficit model and the need for a multivariate approach

As highlighted by McGrath et al. (2020), the multiple deficit model was proposed as a multilevel framework for understanding neurodevelopmental disorders such as ADHD, autism spectrum disorders, and developmental dyslexia. The passage from a single cognitive deficit, conceived as a core deficit, to the more open idea of a set of cognitive deficits that are probabilistically related to the condition labelled as “dyslexia”, would allow one to better explain the high degree of comorbidities between learning disorders. Moreover, the multiple deficit perspective fits better with the empirical evidence provided in the literature about different profiles of readers with dyslexia, characterised by different cognitive deficits (see

Castles & Coltheart, 1993 and the double deficit hypothesis by Wolf & Bowers, 1999). From its first formulation (Pennington, 2006), this multiple deficit account has been tested through several studies that implied multivariate approaches (Moll et al., 2016; Moura et al., 2017; Peterson et al., 2017; Ring & Black, 2018).

The current study attempts to provide a further piece of evidence for supporting the multi-deficit approach and, to this aim, we looked at machine learning as a method that provides a chance to classify children as good and poor readers based on multivariate clinical markers. The adoption of a machine learning–based classification approach has been effective in different disciplines where multivariate clinical markers must be managed: several examples come from biology (see Tarca et al., 2007 and Sommer & Gerlich, 2013 for reviews), medicine (Asri et al., 2016; Ghiasi et al., 2020; Hathaway et al., 2019; Mir & Dhage, 2018), as well as psychology and neuropsychology (e.g., Omar et al., 2019; Rostami et al., 2020; see Battista et al., 2020 and Dwyer et al., 2018 for reviews).

The classification and regression tree (CART) model

CART models are a machine learning technique of modelling to divide and, thus, classify data based on the recursive partition of a given “training” dataset’s feature space (see Myles et al., 2004 for a review). The method, based on multivariate non-parametric correlations, finds a set of decision rules in which input variables are split in root nodes, based on their information gain. Accordingly, the decision tree is built based on the set of sequential rules that better replicate the classification in input. Based on this first training, the CART extracts a series of predictions, i.e., the decision tree model, that can be applied to a new dataset, usually known as the “testing” one (Myles et al., 2004; Pradhan, 2013; Rostami et al., 2020).

This classification approach needs enough variables in input and, thus, is by definition suited to handle a multivariate set of data. The fact itself of obtaining a set of classification rules makes the CART approach a relevant technique to adopt for diagnostic purposes. Moreover, the hierarchical variables’ structure emerging from the tree is highly informative. Indeed, thinking about neuropsychological multiple deficit disorders such as ADHD, autism spectrum, and developmental dyslexia, the use of CART models provides a wide range of advantages (Omar et al., 2019; Rostami et al., 2020). From a diagnostic point of view, these models provide an automated way for classifying participants while also allowing to better understand the role of specific deficits and their reciprocal links. In our study, this approach would provide the opportunity to better understand what language-independent deficits prevail in dyslexia and, based on the number of tree branches, whether we can empirically observe different behavioural profiles. In other words, the adoption of such a multivariate approach allows us to advance in the knowledge of this complex deficit from a comparative perspective, thus supporting from a novel angle the multiple deficit approach proposed by Pennington (2006). Moreover, in line with one of the goals of this work, CART will allow us to apply the decision rules based on several language-independent tasks to MCL, that is to a group for which standardised clinical tests are not yet available.

Aim of the present study

In this study, we aim at validating our *ReadFree* tool by testing criterion validity of the whole screening battery, i.e., a component of construct validity (based on Anastasi, 1986;

Messick, 1979) that represent the degree to which a test can predictively (in the future) or concurrently (in the present) measure something concerning a specific hypothesis (criterion) and, thus, latent construct. To do so, we (*step 1*) first tested the task-by-task criterion validity by testing whether each ReadFree task can be associated with good and poor reading performances. Children were previously identified as good or poor readers by adopting standardised Italian reading tests and clinical criteria (see paragraph 2.1 for a detailed description), i.e., by adopting a set of tasks independent of our *ReadFree* tool to avoid circularity.

Once identified and selected the discriminant tasks, we (*step 2*) explored the between-task patterns of correlations and the nature of their relationship with reading and reading-related cognitive skills. This was done to check whether the ReadFree tasks cover different cognitive-related aspects of the reading process, as they were supposed to do when we developed the screening. In other words, we checked whether the outcomes of our screening can be directly related to reading proficiency, for further testing the validity argument known as “explanation inference”. Indeed, explanation inference is about understanding the theoretical relationship between the test content and the construct of interest, so, in our case, the relationships between ReadFree tasks and the reading process (see Chapelle et al., 2008; Knoch & Chapelle, 2018). This second step was realised through a principal component analysis that let us explore also potential latent factors underlying reading and language-independent cognitive skills. This procedure has been applied only to monolingual good readers to deepen our understanding of aspects related to the typical reading process.

As a third step (*step 3*), we further tested the criterion validity of the whole *ReadFree* tool after excluding non-discriminant tasks. In this step, we adopted a recursive-partitioning machine learning approach based on a classification and regression tree (CART; see paragraph 1.3.1) model to extract a multivariate set of classification rules for discriminating between good and poor monolingual readers. While steps 1 to 3 were focused on the Italian-monolingual sample, in *step 4*, we compared MLC’s performances with those of Italian monolinguals (both good and poor readers) at standardised reading tasks and tasks of the *ReadFree* tool. This step was necessary to check whether we can apply the same reasoning used for monolinguals also to the MLC population. Therefore, we assume that tests capable of identifying poor readers in a population (monolinguals) will be capable of identifying them also in another population with a similar distribution of good and poor readers (MLC), but we have to prove that the two populations have the same reading behaviour and same performances in the ReadFree tasks. This can be considered as a preliminary step to then test the external validity (Campbell, 1957; Ferguson, 2004; see Findley et al., 2021 for a review) of our instrument. In the last step (*step 5*), the classification rules extracted in *step 3* from the sample of monolinguals using the CART model were applied to MLC for identifying poor and good readers in this other group. Once obtained an MLC classification based on the CART predictions, this was compared to the one we obtained by standardised clinical reading tests to compute our tool’s performance measures. Results of these indices are, anyway, to interpret with caution since standardised clinical tests cannot be considered “gold standard” tests for comparison, as we will better explain later in the text. These performance measures will also provide a “reliability” measure of the ReadFree classification rules, as they will represent the goodness of fit of these rules when applied to other participants. Accordingly, with this last step, we will test the generalizability and, thus, the external validity of our tool conceived as the “extent to which inferences drawn from a given study’s sample apply to a broader population or other target populations” (Findley et al., 2021, p. 366).

This multistage validation process (see Anastasi, 1986 for a review) aims at obtaining a final version of the *ReadFree* tool that will be further validated on a larger sample in further studies.

Materials and methods

Participants

A total number of 257 primary and middle school students were involved in the data collection. After having excluded dropouts ($n = 5$), participants with incomplete data on more than two tasks ($n = 3$), participants who had lived for less than a year in Italy, and, as a consequence, were not enough proficient in Italian ($n = 7$), and children with other neurodevelopmental disorders beyond reading difficulties ($n = 14$), we obtained a sample of 228 participants. Of these, 46 were 3rd graders (girls = 24, boys = 22; age in months, mean = 103.83, SD = 4.26), 49 were 4th graders (girls = 22, boys = 27; age in months, mean = 114.94, SD = 4.64), 41 were 5th graders (girls = 29; boys = 12; age in months, mean = 127.8, SD = 3.54), 30 were 6th graders (girls = 15, boys = 15; age in months, mean = 141.2, SD = 3.68), 28 were 7th graders (girls = 12, boys = 16; age in months, mean = 153.82, SD = 5.29), and 34 were 8th graders (girls = 12, boys = 22; age in months, mean = 161.79, SD = 5.53).

Students were included in 3 groups based on their parents' nationality and their performances on reading tests. Accordingly, we obtained (i) a control group of monolingual good readers (GR; $n = 105$) with both Italian parents and without reading disorders; (ii) a group of monolingual students that were evaluated as poor readers (PR; $n = 37$) based on standard clinical reading and cognitive tests; and (iii) a group of minority-language children (MLC; $n = 68$) with one or both foreign parents and a bilingual linguistic family context.

MLC students were heterogeneous for both language of origin and minority-language exposure. The degree of cumulative exposure for each MLC has been investigated using the PLQ Interview (Intervista delle Prassi Linguistiche Quotidiane - Daily Linguistic Practice Interview by Carioti et al., 2022a, b, September 13), a structured interview conceived for assessing language use and experience on both the minority and the majority language on MLC. The amount of time spent speaking the minority language with the mother (average percentage time in a day = 11.23%, SD = 11.7) or father (average percentage time in a day = 5.85%, SD = 6.5) was very heterogeneous and several children ($n = 21$) declared that they did not speak the minority language at home. However, all of them were daily (passively) exposed to the minority language in the family context for at least half an hour (average percentage time in a day = 18.4%, SD = 10.12). Parents' languages of origin, reported in Supplementary Table 1, were very heterogeneous too. All MLC had Italian as the main language of education.

Accordingly, we further excluded children ($n = 18$) with low performance in non-verbal reasoning, i.e., a Raven's matrices' score below the 50th percentile (see paragraph 2.2 for details about the standardised test). Consequently, we obtained a final sample of 210 participants aged between 8 and 13 years old. Demographic information and non-verbal reasoning scores of the three groups are reported in Table 1.

All participants were enrolled in primary and middle public schools in Northern and Central Italy. Data of primary students were collected in the "I.C. Della Torre" of Chiavari (Genova), while data of middle school students were collected in the "I.C. Lanfranco" of Gabicce Mare and Gradara (Pesaro-Urbino). Some other monolingual (15/142 = 10.5%) and MLC (7/68 = 10.2%) participants were enrolled in the Center of Developmental Neuropsychology, AST Pesaro-Urbino. Twenty-two students of the monolingual

Table 1 Demographic data and non-verbal reasoning average scores

	Years	Gender		Handedness (R/L)	N total	Age (months)		Raven's average score (percentile)	
		Girls	Boys			Mean	SD	Mean	SD
<i>GR</i>	8	8	8	16/0	16	102.25	3.51	94.94	4.22
	9	12	13	24/1	25	113.48	3.45	95.08	4.71
	10	15	6	20/1	21	126.9	3.19	94.48	8.57
	11	6	8	14/0	14	139	3.53	80.86	10.63
	12	7	6	12/1	13	150	4.6	83.62	16.03
	13	6	10	16/0	16	161	3.12	86.94	10.13
	Tot.	54	51		105			90.53	10.66
<i>PR</i>	8	2	0	1/1	2	103.5	2.12	94	8.49
	9	6	5	11/0	11	112.82	4.21	83.09	14.78
	10	2	3	4/1	5	124.4	2.97	91.4	15.95
	11	2	3	4/1	5	139.8	4.49	79	10.15
	12	2	5	6/1	7	151.29	1.89	85	18.37
	13	2	5	6/1	7	158.86	2.79	75.86	19.09
	Tot.	16	21		37			83.24	15.77
<i>MLC</i>	8	7	9	15/1	16	102.5	3.06	94.96	6.59
	9	7	11	17/1	18	111.94	3.61	89.39	12.79
	10	9	5	12/2	14	126	3.55	87.58	12.58
	11	6	2	7/1	8	135.62	3.2	92.5	9.62
	12	1	1	2/0	2	154	1.41	86	8.49
	13	5	5	9/1	10	160.6	3.96	80.1	16.67
	Tot.	35	33		68			89.16	12.32

PR group had an official diagnosis of developmental dyslexia (19 in the Center of Developmental Neuropsychology and 3 by a professional neuropsychologist elsewhere). Five more students tested in school were included in the monolingual PR due to their poor reading performances.

None of these children had psychiatric, emotional, or sensory disabilities, and all participants had normal or corrected-to-normal visual acuity. According to the World Medical Association Declaration of Helsinki's ethical principles, informed consent was obtained from parents, and children gave their verbal consent to participate in the study. The Ethical Committee of the University of Urbino Carlo Bo approved the study (prot. Num. 11, 20th August 2018). Some participants (i.e., primary students) were also included in the sample of one of our previous studies (Carioti et al., 2022a, b).

Cognitive assessment

Participants were assessed with the following neuropsychological battery, including standardised clinical reading tests:

- (1) Raven's coloured progressive matrices (CPM; Raven, 1956) and standard progressive matrices (SPM; Raven, 2003; Raven, 1958), that is a set of pattern-matching tasks in which participants must determine the final pattern in a series. This task was used to assess non-verbal reasoning.
- (2) The digit forward and backward subtests of the WISC-IV (Wechsler, 2003), that is a set of tasks that require children to retain in memory a series of digits (forward version) or to retain and reverse the digits' order (backward). These subtests assess both short-term and working memory. Here, the raw score corresponded to the maximum number of digits recalled (i.e., the memory span).
- (3) Nonwords repetition test from the VAUMeLF battery (Batterie per la Valutazione dell'Attenzione Uditiva e della Memoria di Lavoro Fonologica nell'Età Evolutiva; Bertelli & Bilancia, 2008), a task in which 40 nonwords delivered by a recorded voice must be repeated by children. The test assesses auditory attention and phonological skills. If an item was correctly repeated after the first listening, then the child obtained a score of 1; if the nonword was repeated at the second listening, the score was 0.5. The raw score corresponded to the sum of the scores obtained for each item.
- (4) Single word and pseudoword reading was assessed through the DDE-2 test (Batteria per la Valutazione della Dislessia e della Disortografia Evolutiva-2; Sartori et al., 2007), which requires children to read a series of words/pseudowords presented in lists. The word reading test assesses whole-word decoding and, thus, lexical identification, while the pseudoword reading test assesses the phonological decoding skills, i.e., the grapheme-to-phoneme mapping.
- (5) Text reading was assessed using the short-stories in the battery MT and MT-3 Clinica (Cornoldi & Caretti, 2016; Cornoldi & Colpo, 2002).

For each of the 3 reading tasks (words—pseudowords—text), we obtained a fluency (syllables/seconds) and an accuracy score (percentage of accuracy), for a total of 6 measures of reading proficiency.

Experimental tasks

The battery consisted of 12 tasks organised with a hierarchical logic. Half of the tasks were realised in the auditory and the other half in the visual modality to obtain parallel correspondent tasks for both channels (see Table 2 for a summary of the tasks included in the *Read-Free tool*, with detailed description of experimental phases, number of trials for each phase, variants of the task, and the scoring procedure). Only two tasks did not have a correspondent counterpart: the RAN-shapes in the visual modality and the cocktail party effect task in the auditory one. While the first task (the RAN), preceded by a training trial of single naming, was included to test the speed of retrieving a selected pool of lexical labels, the second one (the cocktail party effect task) assessed selective auditory attention. Both were designed to test, even if in a different format, selective attention and to stress “the crowding effect” (see Gori & Facoetti, 2015). Behavioural tasks included in the screening tool are summarised in Table 2. An extended description of each task is included in the supplementary materials.

Experimental procedures

Children were tested individually in two different sessions: in the first one, participants underwent the cognitive battery, and in the second one the *ReadFree* tool was administered.

Table 2 Tasks included in the ReadFree tool

ReadFree tool		Visual modality			
Auditory modality		Visual modality			
<i>Reaction time</i>	<i>Task</i>	<i>N trial</i>	<i>Variants</i>	<i>Scores</i>	
	The participant must click the mouse as soon as a sound is delivered	10		Median RTs of 8 trials (the first two trials are not considered)	
<i>Tapping</i>	Listening: the participant must listen to a rhythmic pulse delivered by a metronome	16	80 bpm 100 bpm		
	Entrainment: the participant must click the mouse aligned to the metronome	16		Tapping onset = degree of anticipation or delay with respect to the target stimulus	
	Free tapping: the participant must reproduce the rhythm without listening to any stimulus	12			
	The participant must click the mouse as soon as a dot appears on the screen	10		Median RTs of 8 trials (The first two trials are not considered)	
	Listening: the participant must watch a dot regularly presented on the screen	16	80 bpm 100 bpm		
	Entrainment: the participant must click the mouse aligned to the dot	16		Tapping onset = degree of anticipation or delay with respect to the target stimulus	
	Free Tapping: the participant must reproduce the rhythm without watching any stimulus	12			

Table 2 (continued)

ReadFree tool		Visual modality					
Auditory modality		Visual modality					
<i>Task</i>	<i>N trial</i>	<i>Variants</i>	<i>Scores</i>	<i>Task</i>	<i>N trials</i>	<i>Variants</i>	<i>Scores</i>
<i>Tone discrimination</i>	The participant must indicate on the keyboard whether the pair of tones presented are “same” or “different”	Adaptive procedure based on Viviani and Stucchi (1989)	Degree of discriminant sensitivity - smallest interval perceived	The participant must indicate on the keyboard whether the pair of grey squares presented are “same” or “different”	Adaptive procedure based on Viviani and Stucchi (1989)		Degree of discriminant sensitivity - smallest interval perceived
<i>Cocktail party</i>	The participant must find a target tone in the noise. The noise is represented by distractors of different timbres and pitch in the range of human voice.	36.5 s of stimulation X 3 trials	d-prime	The participant must rapidly name all the shapes presented in a grid	3	- 7x7 simple matrix - 10x10 matrix - 7x7 matrix with visual interference	Shapes named in 30 seconds

Table 2 (continued)

ReadFree tool		Visual modality			
Auditory modality		Visual modality			
	<i>Task</i>	<i>N trial</i>	<i>Variants</i>	<i>Scores</i>	
<i>Auditory go/no-go</i>	The participant is instructed to click the mouse when a low tone is heard and not to click when a higher one is heard	2	Irregular: random ISI Regular: ISI at 60 bpm	d-prime	
<i>Auditory anticipatory timing</i>	The participant must click the mouse when the second tone of a pair is delivered, aligned with a regular timing	1	Pagliarini et al., 2020	Tapping onset = degree of anticipation or delay compared to the target stimulus	
	The participant is instructed to click the mouse when a grey dot appears and not to click when a yellow one appears	2	Irregular: random ISI Regular: ISI at 60 bpm	d-prime	
	The participant must click the mouse when the second big dot of a pair is delivered, aligned with the regular presentation of smaller dots.	1	<i>Visual anticipatory timing</i>	Tapping onset = degree of anticipation or delay compared to the target stimulus	

All the tasks included in the *ReadFree* tool were developed in the Matlab environment (2018b, www.MathWorks.com) and presented through a PC DELL Inspiron 15 5000, with a 15.6 inches screen, Intel Core™ i7-1165G7 driver, and Windows Home 10 Operative System. Each participant was set in front of the PC and asked to wear headphones (Philips Bass + SHL3075WT/00 with integrated microphone) to listen to the battery's auditory tasks and to provide the vocal answers required in the RAN-shapes task. A Logitech M110 silent mouse was used for participants' responses to avoid any conflicting sound during auditory tasks. Instructions about each task were orally provided by the researchers and some training sessions were included in many tasks (cocktail party, go/no-go, warning imperative). Participants were allowed to ask the researchers for more information before each task started and/or to repeat the training session. No instructions were delivered through text to avoid further complications for poor readers.

Data analyses

All the analyses were performed in the R environment (R Core Team, 2019). Different approaches, implying both unidimensional and multidimensional analyses, were used to test each task's discriminant power, as well as the validity of the *ReadFree* tool.

Step 1

First of all, we explored data distributions through boxplots to detect eventual outliers (i.e., data < 1st or > 3rd quartile + 1.5 IQR). After this check, we used logit models to test the discriminant power of every task; age was included in each model as a covariate. This first step allowed us to reduce the number of tasks involved in the battery and select a set of variables to be included in the multivariate analysis. This technique is commonly used to avoid large standard errors and the risk of identifying spurious associations (Ranganathan et al., 2017). As explained above, this step let us test each task's criterion validity.

Step 2

Performances of monolingual good readers at experimental tasks, reading, and reading-related cognitive skills were included in a PCA. This approach has already been used in several studies to assess the construct validity (particularly, criterion validity) of neuropsychological batteries (Baser & Ruff, 1987; Shum et al., 1990; Vogel et al., 2015), also in the case of computerised tools (Berger et al., 1997; Kabat et al., 2001; Smith et al., 2013). This choice was made because it gave us the chance to explore the pattern of correlations between measures in complex cases, as the present one, and to better explore the relationship between single tasks and specific aspects of a high-order cognitive process underlying accurate and fast reading processing. The analysis was performed using the *principal* function of the "psych" R package (Revelle, 2014) and the *prcomp* function of the "stats" R package. As we assumed a high rate of between-variable correlations, in line with one of our previous studies (Carioti et al., 2019), we applied an Oblimin rotation to factors extracted based on the scree-plot exploration.

Step 3

We adopted a machine learning technique to obtain a hierarchical classification of the discriminant tasks (see paragraph 1.3.1). Accordingly, monolinguals GR and PR data were used to train a classification and regression tree (CART) model (Breiman et al., 1984), by using the *rpart* R package (Therneau et al., 2015). As the classification of GR and PR based on standardised clinical tests was reliable only for Italian monolinguals, only their data were included in the *training dataset*. The number of root nodes included in the decision tree was decided in relation to model's complexity: the tree was cut at the last variable which reduced the complexity. The classification made by the CART model has been compared to the one in input for extracting the ReadFree tool's performance measures. In particular, the performance measures considered were (i) sensitivity, i.e., the number of true positives (TP) on all positive assessments; (ii) specificity, i.e., the number of true negatives (TN) on all true negative assessments; (iii) positive predictive value, i.e., the proportion of positive test results in the group of actual poor readers $[TP/(TP + FP)]$; the negative predictive value, i.e., the proportion of participants that resulted negative for the condition on actual good readers $[TN/(TN + FN)]$; and (iii) the overall accuracy (see Eusebi, 2013; Glaros & Kline, 1988; Trevethan, 2017). The overall accuracy of the instrument, thus calculated as the proportion of true positive (TP) and true negatives (TN) on the entire sample $[(TN + TP)/(TN + TP + FN + FP)]$ (Berlinger et al., 2019; Šimundić, 2009), as the above-mentioned performance indices, will be presented in terms of percentage of participants in which the CART classification corresponds to the one made through standardised clinical reading tests.

Step 4

Performances of Italian monolinguals with and without reading difficulties (monolingual GR + PR) were compared to those of MLC in standardised reading tasks as well as in the discriminant tasks of our *ReadFree* tool, using generalised linear models (GLMs). The group was included into GLMs as a fixed factor and the age (in years) as a covariate. When data did not fit the normal distribution, data transformation and alternative family distributions were applied. When this was not possible, a robust non-parametric model was run, using the *lmrob* of the “robustbase” R package (Maechler et al., 2020) or, in case of censored data, a Tobit regression model was applied (Long, 1997; McDonald & Moffitt, 1980) through the R package VGAM (Yee, 2008). The influence of the socio-economic-status (SES)² on the performance at reading and ReadFree tool's tasks was preliminary checked using the intraclass correlation coefficient (ICC; ICC package—Wolak, 2015) (see Supplementary Table 3). This check let us include in our GLMs only variables that significantly contributed to explaining variance.

As previously mentioned, in *step 4* we wanted to test whether our *ReadFree* tool could be considered empirically suitable for both monolinguals and MLC, assuming that the “language-independent” and “reading free” nature of our tasks make them

² The socio-economical-status (SES) was computed based on the occupation of each child's parents. Occupations were classified using the International Standard Classification of Occupations (ISCO team, International Labour Office) and coded along 10 areas. Based on this classification, mothers and fathers' occupation was collapsed in a unique score, resulting in a three-way classification (high-medium-low level of SES). See Supplementary Table 2 for details on the classification.

effective in identifying poor and good readers in MLC as well as in monolinguals. As previously mentioned (paragraph 1.4), starting from the assumption that the two populations had the same good and poor readers distribution, we would not expect any between-group differences in reading or in the experimental tasks and, accordingly, we could safely assume that the *ReadFree* tool could be employed with MLC for identifying “poor readers”, as it can be with monolinguals. In line with this reasoning, we used generalised linear models (GLMs) to test between-group differences. Here, it is noteworthy that, although between-group differences at every task were tested with a dedicated GLM, the p-values of all GLMs were corrected for false discovery rate (FDR) using the procedure of Benjamini and Yekutieli (2001) to control for multiple comparisons. This step supports our further aim of including MLC in the “testing dataset” of the CART model, extending the prediction obtained from the “training dataset” of monolinguals to a different dataset that can be safely considered as an independent subset of the general dataset.

Step 5

As a final step, the set of rules extracted by the decision tree on Italian monolingual readers (i.e., “training dataset”, see *step 3*) were applied to the data of the MLC group. MLC participants were included in the “Testing dataset” (32.3% of the total sample) and classified by the CART algorithm as poor readers (MLC-PR) or good readers (MLC-GR). The classification of MLC extracted by the CART model was finally compared to the one obtained by applying clinical reading tests to get the *ReadFree* tool’s performance measures presented above. Here again, it is worthy to remember that the normative data of the clinical tests included in the neuropsychological assessment were all collected in Italian-monolingual students. This is something that is not recommended from either a methodological or clinical point of view (see Recommendation 7.3 of the “Italian Guidelines for the Identification of Specific Learning Disorder”). However, the standardised clinical tests were the only set of measures that we could use as reference point to test the performance measures of our *ReadFree* tool.

Results

Step 1. Discriminant power of experimental tasks

The results of the logistic regressions are reported in Table 3. These analyses were run to identify the set of *ReadFree* tool’s tasks capable of discriminating between monolingual GR and PR.

Step 2. Principal component analyses on monolingual GR’s performances

Data of 100 monolingual good readers were included in the PCA. Five participants of the GR group were removed due to missing scores in one *ReadFree* tool’s task.

For each participant, we included in the PCA age (in years), reading and reading-related cognitive tests, and discriminant tasks of the *ReadFree* tool. Patterns of correlations across variables are represented in the heatmap (Fig. 1; the heatmap represents the correlational patterns between reading, cognitive measures, and tasks of the *ReadFree* tool in the sample of monolingual good readers).

Table 3 Results of the logit model run on each task of the ReadFree tool

	GR (N)	PR (N)	Participants (N)	Effect	X ²	DF	p-value
<i>Auditory reaction time</i>	105	36 [^]	141	Task	1.34	1	0.24
				Age	2.81	1	0.09
				Task*Age	0.17	1	0.67
<i>Visual reaction time</i>	101 [^]	35 [^]	136	Task	4.95	1	0.02*
				Age	4.94	1	0.02*
				Task*Age	0.03	1	0.85
<i>Auditory entrainment_{80bpm}</i>	104 [#]	37	141	Task	1.05	1	0.3
				Age	1.39	1	0.23
				Task*Age	0.005	1	0.94
<i>Auditory entrainment_{100bpm}</i>	103 ^{#^}	37	140	Task	4.02	1	0.04*
				Age	3.77	1	0.28
				Task*Age	1.15	1	0.05*
<i>Auditory free tapping_{80bpm}</i>	104 [#]	37	141	Task	6.95	1	.008**
				Age	2.38	1	0.12
				Task*Age	0.08	1	0.76
<i>Auditory free tapping_{100bpm}</i>	103 ^{#^}	37	140	Task	6.54	1	0.01*
				Age	0.45	1	0.5
				Task*Age	1.36	1	0.24
<i>Visual entrainment_{80bpm}</i>	105	37	142	Task	1.06	1	0.3
				Age	1.76	1	0.18
				Task*Age	2.81	1	0.09
<i>Visual entrainment_{100bpm}</i>	104 [^]	37	141	Task	3.43	1	0.09
				Age	1.29	1	0.22
				Task*Age	2.01	1	0.11
<i>Visual free tapping_{80bpm}</i>	105	37	142	Task	1.55	1	0.21
				Age	1.68	1	0.19
				Task*Age	1.55	1	0.21
<i>Visual free tapping_{100bpm}</i>	104 [^]	37	141	Task	3.35	1	0.07
				Age	1.96	1	0.16
				Task*Age	1.87	1	0.17
<i>Tone discrimination</i>	104 [^]	34 [^]	138	Task	0.28	1	0.59
				Age	0.76	1	0.38
				Task*Age	1.81	1	0.17
<i>Grey-scale discrimination</i>	105	37	142	Task	0.46	1	0.49
				Age	1.76	1	0.18
				Task*Age	0.35	1	0.55
<i>Cocktail party task</i>	100 ^{#^}	35 ^{#^}	135	Task	4.43	1	0.05*
				Age	3.56	1	0.11
				Task*Age	3.08	1	0.02*
<i>RAN-shapes</i>	105	37	142	Task	26.38	1	<0.001***
				Age	17.15	1	<0.001***
				Task*Age	0.55	1	0.45

Table 3 (continued)

	GR (N)	PR (N)	Participants (N)	Effect	X ²	DF	p-value
Auditory go/no-go irregular	104 [^]	37	141	Task	3.65	1	<0.001***
				Age	6.04	1	0.013*
				Task*Age	0.00	1	0.99
Auditory go/no-go regular	105	37	142	Task	22.71	1	<0.001***
				Age	7.21	1	0.007**
				Task*Age	3.18	1	0.07
Visual go/no-go irregular	99 ^{#^}	36 [^]	135	Task	5.38	1	0.02*
				Age	2.42	1	0.11
				Task*Age	1.41	1	0.23
Visual go/no-go regular	104 [#]	37	141	Task	13.28	1	<0.001***
				Age	2.05	1	0.15
				Task*Age	0.001	1	0.97
Auditory anticipatory timing	105	37	142	Task	1.08	1	0.29
				Age	1.61	1	0.2
				Task*Age	1.18	1	0.2
Visual anticipatory timing	102 [#]	37	139	Task	1.92	1	0.16
				Age	1.23	1	0.26
				Task*Age	0.02	1	0.86

*** $p < .001$, ** $p < .01$, * $p < .05$

Discriminant tasks are highlighted in bold

[^]Some outliers were removed

[#]Missing data

Based on the scree-plot (see Fig. 2), 4 factors were extracted by the PCA explaining 51% of the variance. Factor loadings produced by the PCA are reported in Table 4 together with commonalities and uniqueness for each variable. Variables with a saturation value $> |0.3|$ were considered to interpret the factors (variables with saturations $> |0.5|$ are reported in bold type).

As clearly emerged from commonalities (see the saturation matrix with loadings, commonalities, and uniqueness in Table 4), the auditory entrainment_{100bpm} and the digit forward had a low contribution to the factorial structure, while the other tasks seemed to be well associated to a specific component of the reading process.

In particular, the first factor (F1) was mainly associated with age, working memory, reading fluency, and RAN (saturation $> |0.5|$). The second factor (F2) mostly represented reading accuracy, which was associated with visual RTs (saturation $> |0.5|$). The third factor (F3) isolated phonological awareness and the entrainment and free tapping, i.e., rhythmic auditory tasks (saturation $> |0.3|$). Lastly, the fourth factor (F4) highlighted the link between the selective auditory attention (cocktail party) and the executive component of inhibition (go/no-go) in both the modalities but did not include any reading or cognitive measures (saturation $> |0.5|$).

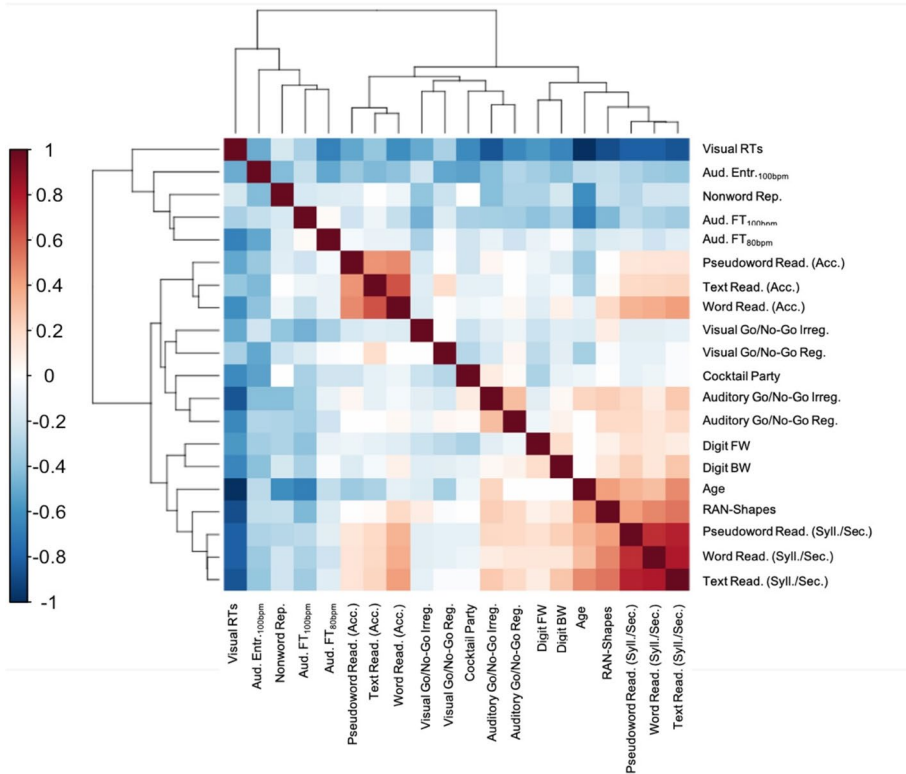


Fig. 1 Heatmap and correlational patterns on monolingual good readers. Negative correlations are depicted in blue, while positive ones in red

Step 3. Classification of good and poor readers based on the CART model

The set of multivariate classification rules is represented in the Decision tree (Fig. 3 and Tables 5–6). Due to some missing data, the CART model was run on the data of 138 participants (102 monolingual GR, 36 monolingual PR). All the discriminant tasks of the *ReadFree* tool, selected with previous univariate analyses, were included as inputs, and the order of root nodes emerged as a function of the information gain score of each task.

Variables identified as root nodes and splitting rules are reported in Table 5, while complexity parameters (cp) of the model are reported in Table 6.

Variable importance order based on the CART model has been reported in Fig. 4. Each variable’s importance value is computed as the sum of the decrease in impurity. This measure is a function of the variable’s role in both “primary splits” and “surrogate splits” (Breiman et al., 1984).

As emerged from the figure and tables, the regular version of the auditory go/no-go, the RAN-shapes, and the auditory entrainment_{100bpm} represented the principal root nodes. The relative cross-validation error for each sub-tree, from smallest to largest, is plotted in Fig. 5. The three is pruned based on the complexity parameters (cp, on the *x*-axis) and on the lowest cross-validation error. The overall performance measures of the CART model

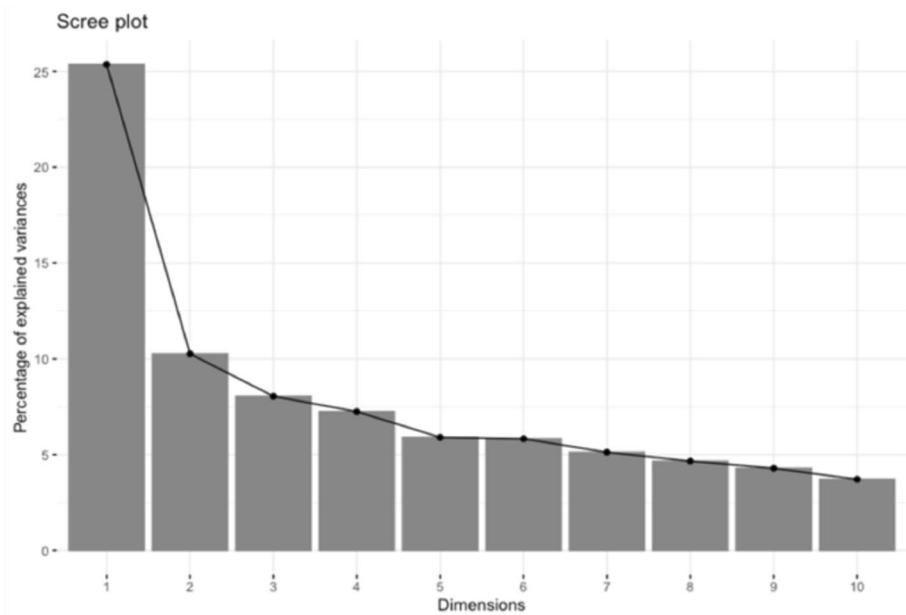


Fig. 2 Scree-plot

Table 4 Saturation matrix

	F1	F2	F3	F4	Commonalities	Uniqueness
<i>Text reading (syll./sec.)</i>	0.89	0.09	0.1	0.02	0.83	0.17
<i>Word reading (syll./sec.)</i>	0.85	0.04	-0.01	-0.12	0.68	0.32
<i>Pseudoword reading (syll./sec.)</i>	0.84	0.13	0.09	0.01	0.76	0.24
<i>Age (years)</i>	0.79	-0.13	-0.16	0.18	0.77	0.23
<i>RAN-shape</i>	0.62	-0.1	-0.13	0.28	0.58	0.42
<i>Digit FW span</i>	0.43	0.13	0.19	-0.03	0.24	0.76
<i>Digit BW span</i>	0.42	0.03	0.27	0.24	0.38	0.62
<i>Text reading (% Acc.)</i>	0.09	0.79	-0.07	0.01	0.65	0.35
<i>Word reading (% Acc.)</i>	0.44	0.63	-0.07	-0.05	0.61	0.39
<i>Pseudoword reading (% Acc.)</i>	-0.04	0.59	0.19	0.08	0.4	0.6
<i>Visual Rts</i>	-0.49	0.56	0.04	0	0.52	0.48
<i>Auditory free tapping_{80bpm}</i>	0.24	-0.03	0.72	0.01	0.57	0.43
<i>Auditory free tapping_{100bpm}</i>	-0.11	-0.05	0.66	0.02	0.44	0.56
<i>Nonword repetition</i>	-0.22	0.14	0.52	0.12	0.34	0.66
<i>Auditory entrainment_{100bpm}</i>	0.07	-0.17	-0.21	0.19	0.12	0.88
<i>Auditory go-nogo (regular)</i>	0.04	-0.07	-0.02	0.74	0.57	0.43
<i>Auditory go-nogo (irregular)</i>	0.3	-0.14	0.13	0.57	0.55	0.45
<i>Visual go-nogo (irregular)</i>	0.01	0.25	-0.38	0.54	0.5	0.5
<i>Visual go-nogo (regular)</i>	-0.12	0.22	0.16	0.53	0.35	0.65
<i>Cocktail party</i>	-0.04	0.04	0.24	0.51	0.31	0.69

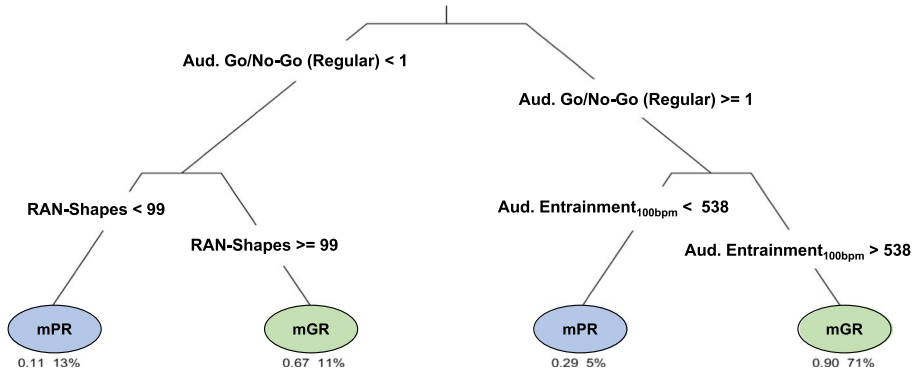


Fig. 3 Root nodes and the classification yielded by the CART model

were obtained through the adoption of a cross-validation procedure. The CART model showed an overall good level of diagnostic accuracy (= 86%, 95% CI: 0.79–0.91), together with a high level of specificity (= 96%, 95% CI: 0.91–0.99) and a low level of sensitivity (= 60%, 95% CI: 0.42–0.75).

Step 4. Comparison between monolinguals and MLC in reading and experimental tasks

In the fourth step of the analysis, the MLC’s performance ($N = 68$) was compared to the one of Italian monolinguals (both poor and good readers; $N = 142$; females = 70, males = 72, age on average = 10.39, SD = 1.67). Firstly, as the SES level was unbalanced between the two groups ($X^2_{(3)} = 26.8, p\text{-value} < .001$), we checked whether SES influenced performances on reading or experimental tasks by computing ICC. The results of this preliminary analysis did not show any significant results (see Supplementary Table 3).

Table 5 Classification rules yielded by the decision tree

	Classification Rules	
	Left son	Right son
<i>Auditory go/no-go (regular)</i>	<1 = PR	≥1 = GR
<i>RAN-shapes</i>	<99 = PR	≥99 = GR
<i>Auditory entrainment_{100bpm}</i>	<538 = PR	>538 = GR

Table 6 Complexity parameters associated to the CART model

Nodes	cp	n splits	Relative error	x error	x standard
<i>Auditory go/no-go (regular)</i>	0.27	0	1	1	0.14
<i>RAN-shapes</i>	0.13	1	0.72	1.02	0.14
<i>Auditory entrainment_{100bpm}</i>	0.08	2	0.59	0.89	0.13
<i>Auditory free tapping_{100bpm}</i>	0.01	3	0.51	0.97	0.14

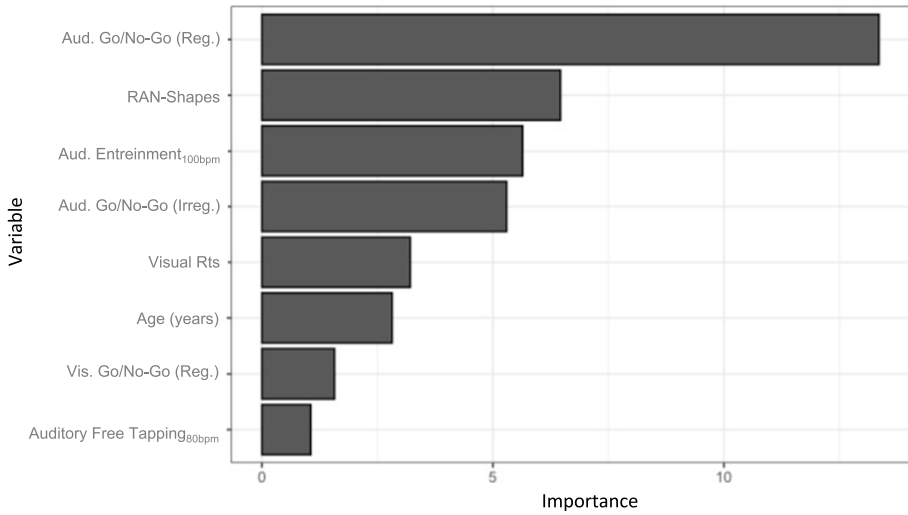


Fig. 4 Variable importance order based on the CART model

The two groups showed similar performances in all reading indices, and they did not differ in any discriminant tasks of the *ReadFree* tool (see Table 7 for results of GLMs testing between-group comparisons on reading speed and accuracy; data distributions in the monolingual Italian sample and in MLC are reported for both reading and experimental tasks in Supplementary Figures 1, 2, 3, 4).

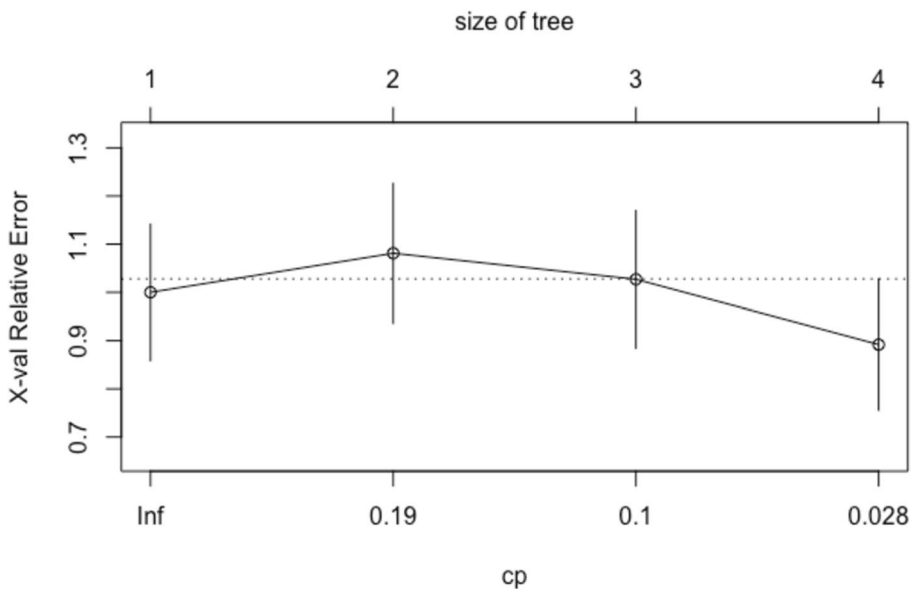


Fig. 5 Relative cross-validation error for each sub-tree

Table 7 Group comparisons on reading skills

Task	Effect	Mono (N)	MLC (N)	Model	Estimate	SE	t	DF	p-value	p-value (FDR)	η^2	R ²
Word reading speed (syll./secs.)	Group	142	68	Linear	0.07	0.14	0.54	207	0.58	1	0.01	0.3*
	Age				0.38	0.04	9.25	207	<.001	<.001	0.28	
Pseudoword reading speed (syll./secs.)	Group	142	68	Linear	-0.06	0.09	-0.63	207	0.52	1	0.001	0.26*
	Age				0.23	0.02	8.5	207	<.001	<.001	0.26	
Text reading speed (syll./secs.)	Group	140 [#]	67 [^]	Linear	0.24	0.16	1.53	204	0.12	0.66	0.007	0.4*
	Age				0.51	0.04	11.37	204	<.001	<.001	0.4	
Word reading accuracy (%)	Group	142	68	Tobit	0.77	0.59	1.29	416	0.19	0.91	--	0.06
	Age				0.64	0.16	3.78	416	<.001*	<.001	--	
Pseudoword reading accuracy (%)	Group	142	68	Tobit	1.13	1.44	0.78	416	0.43	1	--	0.008
	Age				0.72	0.4	1.76	416	0.07	0.57	--	
Text reading accuracy (%)	Group	142	68	Tobit	34	0.38	0.9	416	0.36	1	--	0.005
	Age				0.17	0.1	1.65	416	0.09	0.6		
Auditory free tapping 100bpm	Group	140 [#]	64 [#]	Linear	-0.84	5.46	-0.15	201	0.87	1	0.001	0.01
	Age				-2.77	1.5	-1.84	201	0.06	0.6	0.017	
RAN-shapes	Group	142	68	Linear	-0.71	2.77	-0.25	207	0.79	1	0.004	0.31*
	Age				7.46	0.77	9.65	207	<.001	<.001	0.309	
Auditory go/no-go (irregular)	Group	141 [^]	68	Linear	-0.10	0.11	-0.86	206	0.38	1	0.000	0.16*
	Age				0.29	0.03	6.33	206	<.001	<.001	0.163	
Auditory go/no-go (regular)	Group	141 [^]	68	Linear	-0.13	0.12	-1.11	206	0.26	1	0.001	0.09*
	Age				0.16	0.03	4.66	206	<.001	<.001	0.09	

Table 7 (continued)

Task	Effect	Mono (N)	MLC (N)	Model	Estimate	SE	<i>t</i>	DF	<i>p</i> -value	<i>p</i> -value (FDR)	η^2	R^2
Visual Rts	Group	142	68	Non-param.	-0.003	0.008	-0.39		0.69	1		0.19*
	Age				-0.016	0.002	-7.08		<.001	<.001		
Auditory entrainment _{100bpm}	Group	140 ^{#a}	65 ^{#a}	Non-param.	-1.68	2.6	-0.64		0.51	1		0.004
	Age				0.61	0.72	0.83		0.4	1		
Auditory free tapping _{60bpm}	Group	141 [#]	67 [#]	Non-param.	39.14	23.45	1.66		0.09	0.69		0.03
	Age				10.9	6.48	1.68		0.09	0.69		
Cocktail party	Group	140 [#]	66 [#]	Non-param.	-0.11	0.17	-0.67		0.49	1		0.07
	Age				0.18	0.04	3.85		<.001	0.002		
Visual go/no-go (irregular)	Group	140 ^{#a}	68	Non-param.	-0.15	0.09	-1.69		0.09	0.69		0.05
	Age				0.07	0.02	2.94		0.003	0.04		
Visual go/no-go (regular)	Group	140 ^{#a}	68	Non-param.	-0.07	0.09	-0.75		0.45	1		0.01
	Age				0.03	0.02	1.46		0.14	0.94		

* *F* value associated to the R^2 is significant

In line with these results, we concluded that our screening is suitable for all participants regardless of their linguistic knowledge as the *ReadFree* tool did not penalise, on average, MLC students.

Step 5. Identification of MLC at risk of dyslexia

As a final step, the set of rules extracted from the CART model run on monolinguals (GR, PR) at step 3 was applied to classify the MLC participants (see Table 8 for the confusion matrix including the classification of good and poor readers made through CART model and the clinical tests).

Accordingly, 10 MLC participants were classified as poor readers (MLC-PR) and 58 as good readers (MLC-GR). On the contrary, when clinical criteria based on standard reading tests were applied, 16 MLC were classified as MLC-PR and 52 as MLC-GR. It is noteworthy that only 5 participants classified by the CART as PR corresponded to those identified with clinical criteria (see the confusion matrix in Table 8): the model, thus, returned the 7.3% of false positive rates and the 16.1% of false negative rates, when compared to the clinical classification.

Accordingly, the model showed a low level of sensitivity (= 31%, CI 95%: 0.11–0.59) and a low positive predictive value (= 50%, CI 95%: 0.19–0.81), together with a high level of specificity (= 90%, CI 95%: 0.79–0.97) and high level of negative predictive value (= 81%, CI 95%: 0.69–0.90). The overall accuracy was equal to 76%. Profiles of the 10 MLC participants classified as MLC-PR according to the CART model are reported in Supplementary Table 4. In Supplementary Table 5, we reported the demographic information, parents' nationality, and the severity of reading deficit (i.e., the number of reading parameters < -1.5 ds) for the MLC that underperformed the standardised clinical reading tests.

Discussion

As shown by univariate analyses, 6 tasks out of the 12 included in the original version of the *ReadFree* tool could discriminate between GR and PR. More specifically, visual RTs, auditory entrainment, and free tapping (all conditions except for entrainment_{80bpm}), cocktail party, RAN-shape, auditory go/no-go (both the irregular and regular versions), and visual go/no-go (both the irregular and regular versions) tasks were considered as cognitive markers of reading difficulties in Italian-monolingual students. The CART model applied to the monolinguals accurately identified the most part of participants as good and poor readers (overall accuracy = 86%, CI 95%: 0.79–0.91) and, in particular, it correctly classified as good readers the most part of those monolinguals who resulted as good readers in

Table 8 Confusion matrix

		Clinical Test		
		<i>MLC-PR</i>	<i>MLC-GR</i>	
CART model	<i>MLC-PR</i>	5	5	10
	<i>MLC-GR</i>	11	47	58
		16	52	68

standardised reading tests (specificity = 0.96, CI 95%: 0.91–0.99). Accordingly, the *Read-Free* tool can be considered a promising toolbox for the massive screening of Italian-monolingual students aged between 8 and 13 years, even though we are aware that these indices vary depending on sample numerosity and prevalence of the condition in the sample (see Eusebi, 2013).

These preliminary results set the rationale to “refine” the pool of tests to be included in the *ReadFree* tool and to assess, in a future study, their test–retest reliability and their normative clinical data.

As a matter of fact, the results of the decision tree depicted in Fig. 3 give some intriguing clues on the neuropsychological description of reading difficulties in children.

The tree’s main root, namely the task with the highest discriminant value, is associated with fluid transversal functions, i.e., executive functions as measured by the auditory go/no-go (regular) task. This variable represents the ability to control motor behaviour while extracting regularities (or rhythmic information) from auditory input. Interestingly, this first variable gives rise to two distinct branches: (i) a branch that includes the RAN-shapes task, principally related to automation and integration aspects and (ii) a branch that includes the auditory entrainment in the faster version (100 bpm), associated with the timing component. Thus, our study will be discussed by considering these three cognitive levels and by looking at the relationship between executive functions, automation, auditory timing skills, and the reading process.

Poor reading as a deficit in executive functions, attentional processes, and automation

As mentioned above, the first root node in the decision tree corresponded to the regular version of the auditory go/no-go task; namely, the go/no-go task in which stimuli are delivered with a rhythmic interval. This result is particularly relevant because it suggests that managing regularity and inhibition was highly demanding for the PR group. Indeed, this version of the task requires (i) to update a continuous auditory flow of information, (ii) to process the timing organisation of the stimuli, (iii) to discriminate between the “Go” and the “No-Go” auditory signals. Nevertheless, this specific result seems to better reflect the role of executive functions rather than the role of regularity processing itself. Indeed, one mandatory condition for performing the task well is to inhibit the tendency to “Go”, that can be increased due to the regular presentation of stimuli. Accordingly, we suggest that getting ready to react based on a rhythmic pulse, even when not required by the type of stimulus (namely in the “No-Go” condition) can produce a higher number of “False Alarms”. Therefore, due to the regularity, the stimulus’ arrival is predictable (Large & Jones, 1999; Pagliarini et al., 2020) and the tendency to react at a specific time needs to be efficiently inhibited for avoiding an incorrect response. This may suggest that a good rhythmic awareness could represent a disadvantage. In other words, because of the regularity, executive functions would be the cognitive aspect more stressed by this task and, based on our results, the most powerful language-independent marker of reading difficulties.

The role of executive functions in reading disorders is further supported, based on univariate analyses, by the fact that also the irregular go/no-go task discriminated between monolingual good and poor readers, both in the auditory and visual modality. The crucial role of executive functions, and particularly inhibition, in reading and dyslexia has been deeply explored in the last 20 years (see the review by Farah et al., 2021),

in children (e.g., Doyle et al., 2018; Moura et al., 2014; Reiter et al., 2005; Varvara et al., 2014), as well as in adults (e.g., Brosnan et al., 2002; Smith-Spark et al., 2016). For example, Varvara et al. (2014) suggested that dyslexia is characterised by a global deficit in higher-order domain-general cognitive mechanisms, namely a deficit of executive control regardless of the specific modality of stimuli presentation (visual or auditory). However, one may argue that executive functions impairment is a ubiquitous condition across neurodevelopmental disorders, something that can give clinicians some clue about the presence of a neurodevelopmental condition (a general sign) without representing any specific pathology (Willcutt et al., 2008). This line of reasoning is further supported by the results of the logistic regressions on monolingual children. We found that the visual RT task discriminated between monolingual GR and PR: this result supports both the idea of a deficit in attentional orientation and focusing, as already suggested by Facoetti, Paganoni, and Lorusso (2000a), and the possibility of a deficit at the lower cognitive level of alerting (Facoetti et al., 2019; Posner & Petersen, 1990). Once again, this would not be specific to reading deficits but would instead represent a generic sign shared with other neurodevelopmental conditions such as ADHD (Mullane et al., 2011). The deficit in executive functions and attentional components would also be domain independent. For example, we found significant differences between monolingual GR and PR also at the cocktail party task, i.e., in the auditory attentional domain. The performance at this task implies, by definition, the inhibition of auditorily-presented noise distractors and auditory selective attention.

It is noteworthy that, if one considers the PCA run on the monolingual GR group, executive functions represent a factor that has an independent contribution with respect to reading and phonological skills (see F4 in Table 4). This factor included both inhibition tasks, as the go/no-go in all its versions and modalities, and the selective attention measured through the cocktail party task.

Taken together, the results of the decision tree's root, those of univariate analyses (i.e., the logistic regressions) and of the PCA seem to support the idea that a large part of the poor readers might be characterized by poor executive functions and attentional deficits in both the visual and auditory domain.

However, more specific cognitive signs of reading difficulties emerge if one climbs the decision tree from this general root (i.e., executive functions and attention). In particular, the second node of the decision tree is represented by the performance at the RAN-shapes. Here, it is noteworthy that children who performed worse in the auditory go/no-go, must have also shown low naming skills at RAN to be classified as poor readers according to the results of our CART model (Fig. 3). This result is in line with many studies suggesting that RAN skills are one of the most reliable markers of dyslexia across ages and orthographies (see Araújo & Faísca, 2019; Carioti et al., 2021; Norton & Wolf, 2012 for reviews). Moreover, by adopting a non-alphanumeric novel version of the task, which minimised the role of the reading system (see Carioti et al., 2022a, b for further details), we were able to expand further the empirical findings that support the double deficit hypothesis (Wolf & Bowers, 1999). Interestingly, the association between the deficit at the auditory go/no-go (regular) and RAN tasks in the left branch of the decision tree suggests that the ability to inhibit concomitant distractors, concerning visual stimulus and lexical labels, might represent the core link in the RAN-reading relationship (Bexkens et al., 2015; Van Reybroeck & De Rom, 2019). An ad hoc created experimental paradigm should better address this hypothesis. Interestingly, the results of our PCA analysis further support the RAN-reading relationship. In particular, monolingual GRs

showed a strong relationship between RAN and reading fluency (Table 4, F1), in line with several authors (Lervåg & Hulme, 2009; Papadopoulos et al., 2016, for reviews see Norton & Wolf, 2012 and Araújo & Faísca, 2019). Interestingly, the RAN measure did not contribute to the second factor, namely reading accuracy. These findings support the idea that RAN is more deeply related to reading fluency than to reading accuracy and, thus, to the aspect that more clearly reveals the acquired automation of the reading process.

As a final remark, we would like to stress that the dissociation between accuracy and fluency measures in reading is in line with the results reported in one of our previous studies (Carioti et al., 2019) on a completely different sample of participants.

The other side of poor reading: a deficit of timing skills

Moving to the right branch of the decision tree, the auditory timing deficit seems to prevail. The ability to tap in entrainment with a metronome (entrainment_{100bpm}) was particularly relevant for the classification of those poor readers that can adequately perform the regular auditory version of the go/no-go task. The idea of a deficit in readers with dyslexia in conceiving, organising, and reproducing a simple regular pulsation, i.e., a regular pattern, supports causal theories of dyslexia concerning the perception of rapid spectro-temporal alterations of auditory stimuli (see the RAP theory by Tallal, 1980; see Tallal & Gaab, 2006), as well as temporal awareness of regular strong-weak patterns (see the temporal sampling hypothesis by Goswami et al., 2011; Huss et al., 2011). This can be said also for what concerns the cerebellar theory (Nicolson et al., 2001b), if one considers the rhythmic motor production, and the ensuing implicit learning deficits (Gabay et al., 2015; Kahta & Schiff, 2016; Nigro et al., 2015).

According to these theoretical perspectives, timing skills would underpin phonological awareness and produce a cascade influence on reading (Bishop & Snowling, 2004; Démonet et al., 2004; Gabrieli, 2009; Pagliarini et al., 2020; Peterson & Pennington, 2015; Vellutino et al., 2004). This claim seems further supported by the results of the PCA on monolingual good readers. Indeed, rhythmical tasks such as entrainment and free tapping represented a third independent factor together with nonword repetition, which is the main phonological task included in our cognitive assessment. Our results, thus, suggest that timing skills may contribute to identifying poor readers beyond executive functions and automation deficit.

Since a task that correlated with phonology (the entrainment task) represented an independent decision tree's root, dissociated to the RAN's one, our results seem, once again, seem to support the double deficit hypothesis (Wolf & Bowers, 1999).

The fact that different rules, depending on different tasks' scores, can be applied to recognise a poor reader, would also support the view of the multifactorial deficit account (Peterson & Pennington, 2015; for a critical revision, see Compton, 2021). In this perspective, timing skills might represent another relevant language-independent aspect underlying the reading processes.

A *ReadFree* tool for the assessment of minority-language children

We started the cross-cultural validation of the *ReadFree* tool by a simple basic assumption: the prevalence of reading deficits due to neurodevelopmental conditions should

be similar in an unselected group of monolinguals (i.e., in a sample including both good and poor readers) and in a sample of MLC. We therefore carried out a preliminary between-group comparison and the empirical results showed that, on average, the two groups performed similarly both on the standardised clinical reading tests and on the experimental tasks included in the reduced version of the *ReadFree* tool.

Accordingly, we applied the classification rules extracted from the decision tree trained on the monolinguals to the MLC group. As reported in the results section, in the MLC group, the CART model's automated classification identified 10 participants out of 68 MLC as PR (i.e., 14.7% of the sample). However, when we instead applied the clinical Italian standard criteria to MLC, 16 children out of 68 showed reading difficulties (i.e., 23.5%), a proportion well above the one identified by the *ReadFree* tool. This result suggests that the adoption of Italian monolinguals' normative data might introduce a bias if applied on MLC; an issue that has been recently highlighted also in the new "Italian guidelines for the identification of Specific Learning Disorder" (see the Recommendation 7.3, cited in the introduction). This may be also the reason why we obtained a low level of sensitivity in our method comparison (sensitivity = 31%; 95% CI: 0.11–0.59). Here, it is noteworthy that the sensitivity has not been computed against a "gold-standard" measure, rather, against a set of psychometrics measures that, by definition, are referred to a population different from MLC, i.e., to monolingual children. Consequently, the standardised Italian reading tests do not seem to be an adequate gold standard to compute reliable indices of sensitivity and specificity (see Trevethan, 2017 for more details) and, thus, they must be carefully interpreted when considering the classification applied on MLC.

In line with the issue of assessing bilinguals for language and reading skills, recently, in Italy, attempts to develop a battery to assess verbal and non-verbal competencies in bilinguals have been pursued (BaBIL; Contento et al., 2013; Eikerling et al., 2022). Similarly, Marinelli et al. (2020) provided normative data for bilingual students on a vast range of neuropsychological tests, including reading. However, these clinical instruments are not available for large-scale screening assessments in the school. The *ReadFree* tool represents a useful tool to fill this gap of knowledge and practice; it represents an easy to manage tool for teachers and clinicians that can be applied in the school irrespectively by the children's linguistic background. We could say that it is the first form of an *inclusive screening* tool in Italy. Moreover, its language-independent nature makes it easy to be also transposed in other countries.

Open issues and future directions

Although our *ReadFree* tool, when applied on MLC, obtained an adequate level of performance expressed in terms of overall accuracy (0.76) and a high level of specificity (0.9), if one looks at the confusion matrix reported in Table 8, some intriguing evidence emerged. First, there are 5 MLC classified as PRs by the CART model that, however, did not show any difficulty in the clinical reading tests. These could be considered "False-positive cases" that emerged due to a failure of the CART model classification or due to some neuropsychological factors that are worth a discussion.

One possible (and optimistic) hypothesis is that our screening tool could identify children that will manifest—but do not already show—a reading deficit. This was likely the case for only one child out of 5, given that this child (20ADF, see Supplementary Table 4) showed a failure in one reading parameter (word reading

accuracy, see Supplementary Table 4). Further longitudinal studies will be, in this context, very useful to clarify this issue and, at the same time, to understand whether the *ReadFree* tool can predict future manifestations of reading deficits also in younger children.

Another scenario, probably more adherent to our results, is related to the type of deficit recognised by our tool. Although we developed the *ReadFree* tool to assess reading deficits, the monolingual PRs included in the training dataset were heterogeneous, as they included children with a certified clinical diagnosis and children with subtler reading deficits that did not receive yet a formal diagnosis. This view is further supported by some clinical considerations: the diagnosis more often reported in the monolingual PR group is “General Learning Disorders” (12 out of 22 certified participants). This suggests that children included in our sample reported a reading deficit in comorbidity with other learning disorders. Even though this may represent a methodological limitation of our study, it represents an advantage from an ecological point of view: the high level of comorbidities and shared cognitive deficit between dyslexia and dyscalculia (Cheng et al., 2018; Peters et al., 2018; Peters et al., 2020), dyslexia and writing disorders (Döhla & Heim, 2015; Ehri, 2000; Richards et al., 2015) and, more in general, dyslexia and other neurodevelopmental disorders is widely documented (for a review, see Hendren et al., 2018). This fact induced the authors of the latest version of the Diagnostic and Statistical Manual of Mental Disorder (DSM-5; American Psychiatric Association, 2013) to group all learning disabilities under the more comprehensive definition of specific learning disorders (SLDs) and, as highlighted by Hendren et al. (2018, p. 4), “the subtypes of SLDs have been viewed from an academic-subject approach”. Accordingly, a second validation study in which the assessment of children with SLDs is coupled with a more comprehensive neuropsychological assessment (including, for example, arithmetic and writing skills) is needed to better address this issue. Moreover, in order to make our *ReadFree* tool a standard screening practice, we would need to test another independent and broader sample to assess the replicability of the decision tree, test–retest reliability, and extract normative data of the *ReadFree* tool’s final version.

Furthermore, based on diagnostic parameters, sensitivity levels of our screening were poor in both monolinguals (sensitivity = 60%) and MLC (sensitivity = 31%). Thus, regardless of the issue concerning standardised reading tests applied to MLC, our tool seems to be limited in its ability to detect poor readers. In other words, only 60% of children are correctly recognised as poor readers by our tool, while 40% of them will be classified as good readers. Although this would represent a limit of our tool, we think the sensitivity of our screening will be improved by increasing the sample in further studies and by better studying the functional neuropsychological profile of deficits through an extensive neuropsychological battery. Anyway, by looking at the encouraging good levels of specificity (= 90%) and accuracy (= 88%) in monolinguals and considering that the tool has been developed as a first exploratory screening to orient decisions about the need for subsequent clinical evaluations, we believe that the *ReadFree* may constitute a promising tool that could be very useful if adopted in a school setting.

To conclude, auditory regular go/no-go task, RAN-shapes, and auditory entrainment_{100bpm} emerged as good predictors of reading skills, supporting the idea that some specific integrative and inter-related cognitive components concerning executive functions, attention, and timing are at the basis of the reading process, beyond phonological processing. These results are in line with the universal neuropsychological markers

highlighted in the recent meta-analytic study by Carioti et al. (2021) and with the multi-factorial view of developmental dyslexia (McGrath et al., 2020; Pennington, 2006).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11881-023-00287-3>.

Acknowledgments Part of this work has been realised for the Ph.D. project of the first author, who wants to thank for the support her professors, colleagues, roommates, (old and new) friends, and family. Special thanks to teachers and headteachers of the “I.C. Della Torre” (Chiavari, Genova, Italy) and “I.C. Lanfranco” (Gabicce Mare-Gradara, Pesaro-Urbino, Italy), especially to Daniela Gentilucci and Giuseppe Polverari, and to all students that participated in the study. Finally, thanks to Susanna Laurenzi and Gianmarco Giovagnoli for helping us in scoring and data elaboration.

Author contribution DC, MB, NS, and MTG contributed to the study’s conception and design. CT programmed the ReadFree’s tasks in the Matlab environment. DC, MFM, SS collected experimental data; DC, MDM, SS, ST performed the cognitive assessment. DC, supervised by MB, performed statistical analyses. DC wrote the first version of the draft, while MB, MV, MT, and MTG revised it. All authors commented on previous versions of the manuscript and approved the final manuscript.

Funding Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement. This work was supported by an agreement between ASUR Marche, Area Vasta 1, Pesaro and the Department of Humanities, University of Urbino Carlo Bo.

Data Availability Data will be available on request to the corresponding author.

Declarations

Conflict of interest The authors declare no competing interests.

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

References

- Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology*, 37(1), 1–16.
- American Psychiatric Association, D. S. M. T. F., & American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders: DSM-5 (Vol. 5, No. 5). Washington, DC: American psychiatric association.
- Araújo, S., & Faisca, L. (2019). A meta-analytic review of naming-speed deficits in developmental dyslexia. *Scientific Studies of Reading*, 23(5), 349–368. <https://doi.org/10.1080/10888438.2019.1572758>
- Arikan, S., van de Vijver, F. J., & Yagmur, K. (2017). PISA mathematics and reading performance differences of mainstream European and Turkish immigrant students. *Educational Assessment, Evaluation and Accountability*, 29(3), 229–246. <https://doi.org/10.1007/s11092-017-9260-6>
- Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, 83, 1064–1069. <https://doi.org/10.1016/j.procs.2016.04.224>

- Azzolini, D., Schnell, P., & Palmer, J. (2012). Educational achievement gaps between immigrant and native students in two “new immigration countries”: Italy and Spain in comparison. *The Annals of the American Academy of Political and Social Science*, 643(1), 46–77. <https://doi.org/10.1177/0002716212441590>
- Banai, K., & Ahissar, M. (2010). On the importance of anchoring and the consequences of its impairment in dyslexia. *Dyslexia*, 16(3), 240–257. <https://doi.org/10.1002/dys.407>
- Baldeweg, T., Richardson, A., Watkins, S., Foale, C., & Gruzelier, J. (1999). Impaired auditory frequency discrimination in dyslexia detected with mismatch evoked potentials. *Annals of Neurology*, 45(4), 495–503. [https://doi.org/10.1002/1531-8249\(199904\)45:4<495::AID-ANA11>3.0.CO;2-M](https://doi.org/10.1002/1531-8249(199904)45:4<495::AID-ANA11>3.0.CO;2-M)
- Barbosa, T., Rodrigues, C. C., Mello, C. B. D., Silva, M. C. D. S., & Bueno, O. F. A. (2019). Executive functions in children with dyslexia. *Arquivos de Neuro-Psiquiatria*, 77, 254–259. <https://doi.org/10.1590/0004-282X20190033>
- Baser, C. A., & Ruff, R. M. (1987). Construct validity of the San Diego neuropsychological test battery. *Archives of Clinical Neuropsychology*, 2(1), 13–32. <https://doi.org/10.1093/arclin/2.1.13>
- Battista, P., Salvatore, C., Berlingeri, M., Cerasa, A., & Castiglioni, I. (2020). Artificial intelligence and neuropsychological measures: The case of Alzheimer’s disease. *Neuroscience & Biobehavioral Reviews*, 114, 211–228. <https://doi.org/10.1016/j.neubiorev.2020.04.026>
- Benassi, M., Simonelli, L., Giovagnoli, S., & Bolzani, R. (2010). Coherence motion perception in developmental dyslexia: A meta-analysis of behavioral studies. *Dyslexia*, 16(4), 341–357. <https://doi.org/10.1002/dys.412>
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29, 1165–1188. <https://doi.org/10.1214/aos/1013699998>
- Berger, S. G., Chibnall, J. T., & Gfeller, J. D. (1997). Construct validity of the computerized version of the Category Test. *Journal of Clinical Psychology*, 53(7), 723–726. [https://doi.org/10.1002/\(sici\)1097-4679\(199711\)53:7<723::aid-jclp9>3.0.co;2-i](https://doi.org/10.1002/(sici)1097-4679(199711)53:7<723::aid-jclp9>3.0.co;2-i)
- Berlingeri, M., Devoto, F., Gasparini, F., Saibene, A., Corchs, S. E., Clemente, L., Danelli, L., Gallucci, M., Borgoni, R., Borghese, N. A., & Paulesu, E. (2019). Clustering the brain with “CluB”: A new toolbox for quantitative meta-analysis of neuroimaging data. *Frontiers in Neuroscience*, 13, 1037. <https://doi.org/10.3389/fnins.2019.01037>
- Bertelli B. Bilancia G. (2008), VAU-MeLF: Batteria per la valutazione dell’attenzione uditiva e della memoria di lavoro fonologica nell’età evolutiva, Firenze, Giunti OS
- Bexkens, A., van den Wildenberg, W. P., & Tijms, J. (2015). Rapid automatized naming in children with dyslexia: Is inhibitory control involved? *Dyslexia*, 21(3), 212–234. <https://doi.org/10.1002/dys.1487>
- Bishop, D. V., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*, 130(6), 858–886. <https://doi.org/10.1037/0033-2909.130.6.858>
- Boll-Avetisyan, N., Bhatara, A., & Höhle, B. (2020). Processing of rhythm in speech and music in adult dyslexia. *Brain Sciences*, 10(5), 261. <https://doi.org/10.3390/brainsci10050261>
- Bonacina, S., Cancer, A., Lanzi, P. L., Lorusso, M. L., & Antonietti, A. (2015). Improving reading skills in students with dyslexia: The efficacy of a sublexical training with rhythmic background. *Frontiers in Psychology*, 6, 1510. <https://doi.org/10.3389/fpsyg.2015.01510>
- Bonifacci, P., & Tobia, V. (2016). Crossing barriers: Profiles of reading and comprehension skills in early and late bilinguals, poor comprehenders, reading impaired, and typically developing children. *Learning and Individual Differences*, 47, 17–26. <https://doi.org/10.1016/j.lindif.2015.12.013>
- Booth, J. N., Boyle, J. M., & Kelly, S. W. (2010). Do tasks make a difference? Accounting for heterogeneity of performance of children with reading difficulties on tasks of executive function: Findings from a meta-analysis. *The British Journal of Developmental Psychology*, 28(Pt 1), 133–176. <https://doi.org/10.1348/026151009x485432>
- Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature*, 271, 746–747.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC press.
- Brosnan, M., Demetre, J., Hamill, S., Robson, K., Shepherd, H., & Cody, G. (2002). Executive functioning in adults and children with developmental dyslexia. *Neuropsychologia*, 40(12), 2144–2155. [https://doi.org/10.1016/s0028-3932\(02\)00046-5](https://doi.org/10.1016/s0028-3932(02)00046-5)
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54, 297–312. <https://doi.org/10.1037/h0040950>
- Cancer, A., Bonacina, S., Antonietti, A., Salandi, A., Molteni, M., & Lorusso, M. L. (2020). The effectiveness of interventions for developmental dyslexia: Rhythmic reading training compared with

- hemisphere-specific stimulation and action video games. *Frontiers in Psychology*, *11*, 1158. <https://doi.org/10.3389/fpsyg.2020.01158>
- Carioti, D., Danelli, L., Guasti, M. T., Gallucci, M., Perugini, M., Steca, P., ... & Paulesu, E. (2019). Music education at school: Too little and too late? Evidence from a longitudinal study on music training in preadolescents. *Frontiers in Psychology*, *10*, 2704. <https://doi.org/10.3389/fpsyg.2019.02704>
- Carioti, D., Masia, M. F., Travellini, S., & Berlingeri, M. (2021). Orthographic depth and developmental dyslexia: A meta-analytic study. *Annals of Dyslexia*, *1-40*. <https://doi.org/10.1007/s11881-021-00226-0>
- Carioti, D., Stucchi, N. A., Toneatto, C., Masia, M. F., Broccoli, M., Carbonari, C., Travellini, S., Del Monte, M., Riccioni, R., Marcelli, A., Vernice, M., Guasti, M. T., & Berlingeri, M. (2022a). RAN as a universal marker of developmental dyslexia in Italian monolingual and minority-language children. *Frontiers in Psychology*, *13*, 783775. <https://doi.org/10.3389/fpsyg.2022.783775>
- Carioti, D., Stefanelli, S., Masia, M. F., Giorgi, A., Del Pivo, G., Del Monte, M., ... Berlingeri, M. (2022b). The daily linguistic practice interview: A new instrument to assess language use and experience in minority-language children and their effect on reading skills. <https://doi.org/10.31234/osf.io/et4hx>
- Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, *47*(2), 149–180. [https://doi.org/10.1016/0010-0277\(93\)90003-E](https://doi.org/10.1016/0010-0277(93)90003-E)
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). *Building a validity argument for the test of english as a foreign language*. Routledge; Taylor & Francis Group.
- Cheng, D., Xiao, Q., Chen, Q., Cui, J., & Zhou, X. (2018). Dyslexia and dyscalculia are characterized by common visual perception deficits. *Developmental Neuropsychology*, *43*(6), 497–507. <https://doi.org/10.1080/87565641.2018.1481068>
- Compton, D. L. (2021). Focusing our view of dyslexia through a multifactorial lens: A commentary. *Learning Disability Quarterly*, *44*(3), 225–230. <https://doi.org/10.1177/0731948720939009>
- Contento, S., Bellochi, S., Bonifacci, P. (2013), BaBIL. Prove per la valutazione delle competenze verbali e non verbali in bambini bilingui, Giunti O.S, Firenze.
- Cornoldi, C., & Colpo, G. (2002). *Nuove prove di lettura MT per la scuola secondaria di I grado*. O.S. Organizzazioni Speciali.
- Cornoldi, C., & Caretti, B. (2016). *Prove MT 3 Clinica- scuola primaria e secondaria di I grado*. Giunti Psychometrics.
- Danelli, L., Berlingeri, M., Bottini, G., Borghese, N. A., Lucchese, M., Sberna, M., Price, C. J., & Paulesu, E. (2017). How many deficits in the same dyslexic brains? A behavioural and fMRI assessment of comorbidity in adult dyslexics. *Cortex*, *97*, 125–142. <https://doi.org/10.1016/j.cortex.2017.08.038>
- Démonet, J. F., Taylor, M. J., & Chaix, Y. (2004). Developmental dyslexia. *Lancet*, *363*(9419), 1451–1460. [https://doi.org/10.1016/S0140-6736\(04\)16106-0](https://doi.org/10.1016/S0140-6736(04)16106-0)
- Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized' naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*(4), 471–479. [https://doi.org/10.1016/0028-3932\(76\)90075-0](https://doi.org/10.1016/0028-3932(76)90075-0)
- Döhla, D., & Heim, S. (2015). Developmental dyslexia and dysgraphia: What can we learn from the one about the other? *Frontiers in Psychology*, *6*, 2045. <https://doi.org/10.3389/fpsyg.2015.02045>
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431. [https://doi.org/10.1016/0001-6918\(69\)90065-1](https://doi.org/10.1016/0001-6918(69)90065-1)
- Doyle, C., Smeaton, A. F., Roche, R. A. P., & Boran, L. (2018). Inhibition and updating, but not switching, predict developmental dyslexia and individual variation in reading Ability. *Frontiers in Psychology*, *9*, 795. <https://doi.org/10.3389/fpsyg.2018.00795>
- Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*, 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>
- Eikerling, M., Secco, M., Marchesi, G., Guasti, M. T., Vona, F., Garzotto, F., & Lorusso, M. L. (2022). Remote dyslexia screening for bilingual children. *Multimodal Technologies and Interaction*, *6*(1), 7. <https://doi.org/10.3390/mti6010007>
- Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders*. <https://doi.org/10.1097/00011363-200020030-00005>
- Elbro, C., & Jensen, M. N. (2005). Quality of phonological representations, verbal learning, and phoneme awareness in dyslexic and normal readers. *Scandinavian Journal of Psychology*, *46*(4), 375–384. <https://doi.org/10.1111/j.1467-9450.2005.00468.x>

- Eusebi, P. (2013). Diagnostic accuracy measures. *Cerebrovascular Diseases*, 36(4), 267–272. <https://doi.org/10.1159/000353863>
- Everatt, J., Smythe, I., Adams, E., & Ocampo, D. (2000). Dyslexia screening measures and bilingualism. *Dyslexia*, 6(1), 42–56. [https://doi.org/10.1002/\(SICI\)1099-0909\(200001/03\)6:1<42::AID-DYS157>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1099-0909(200001/03)6:1<42::AID-DYS157>3.0.CO;2-0)
- Facoetti, A., & Molteni, M. (2001). The gradient of visual attention in developmental dyslexia. *Neuropsychologia*, 39(4), 352–357. [https://doi.org/10.1016/s0028-3932\(00\)00138-x](https://doi.org/10.1016/s0028-3932(00)00138-x)
- Facoetti, A., Franceschini, S., & Gori, S. (2019). Role of visual attention in developmental dyslexia. In Verhoeven, Ludo., Perfetti, Charles., Pugh, Kenneth (Eds.), *Developmental Dyslexia Across Languages and Writing Systems*, 307–326.
- Facoetti, A., Paganoni, P., & Lorusso, M. L. (2000a). The spatial distribution of visual attention in developmental dyslexia. *Experimental Brain Research*, 132(4), 531–538. <https://doi.org/10.1007/s002219900330>
- Facoetti, A., Paganoni, P., Turatto, M., Marzola, V., & Mascetti, G. G. (2000b). Visual-spatial attention in developmental dyslexia. *Cortex*, 36(1), 109–123. [https://doi.org/10.1016/s0010-9452\(08\)70840-2.10.3758/BF03210983](https://doi.org/10.1016/s0010-9452(08)70840-2.10.3758/BF03210983)
- Farah, R., Ionta, S., Horowitz-Kraus, T. (2021). Neuro-behavioral correlates of executive dysfunctions in dyslexia over development from childhood to adulthood. *Frontiers in Psychology*, 12, 3236. <https://doi.org/10.3389/fpsyg.2021.708863>
- Farmer, M. E., & Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin & Review*, 2(4), 460–493.
- Ferguson, L. (2004). External validity, generalizability, and knowledge utilization. *Journal of Nursing Scholarship*, 36(1), 16–22. <https://doi.org/10.1111/j.1547-5069.2004.04006.x>
- Findley, M. G., Kikuta, K., & Denly, M. (2021). External validity. *Annual Review of Political Science*, 24, 365–393. <https://doi.org/10.1146/annurev-polisci-041719-102556>
- Flaunacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS One*, 10(9), e0138715. <https://doi.org/10.1371/journal.pone.0138715>
- Flaunacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S., Monasta, L., Montico, M., Sila, A., Ronfani, L., & Schön, D. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in Human Neuroscience*, 8, 392. <https://doi.org/10.3389/fnhum.2014.00392>
- Fostick, L., Bar-El, S., & Ram-Tsur, R. (2012). Auditory temporal processing and working memory: Two independent deficits for dyslexia. *Online Submission*, 2(5), 308–318.
- Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K., & Facoetti, A. (2012). A causal link between visual spatial attention and reading acquisition. *Current Biology*, 22(9), 814–819. <https://doi.org/10.1016/j.cub.2012.03.013>
- Gaab, N., Gabrieli, J. D., Deutsch, G. K., Tallal, P., & Temple, E. (2007). Neural correlates of rapid auditory processing are disrupted in children with developmental dyslexia and ameliorated with training: an fMRI study. *Restorative Neurology and Neuroscience*, 25(3–4), 295–310.
- Gabay, Y., Thiessen, E. D., & Holt, L. L. (2015). Impaired statistical learning in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*, 58(3), 934–945. https://doi.org/10.1044/2015_JSLHR-L-14-0324
- Gabrieli, J. D. (2009). Dyslexia: a new synergy between education and cognitive neuroscience. *Science*, 325(5938), 280–283. <https://doi.org/10.1126/science.1171999>
- Georgiou, G. K., Protopapas, A., Papadopoulos, T. C., Skaloumbakas, C., & Parrila, R. (2010). Auditory temporal processing and dyslexia in an orthographically consistent language. *Cortex*, 46(10), 1330–1344. <https://doi.org/10.1016/j.cortex.2010.06.006>
- Ghiassi, M. M., Zendejboudi, S., & Mohsenipour, A. A. (2020). Decision tree-based diagnosis of coronary artery disease: CART model. *Computer Methods and Programs in Biomedicine*, 192, 105400. <https://doi.org/10.1016/j.cmpb.2020.105400>
- Glaros, A. G., & Kline, R. B. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. *Journal of Clinical Psychology*, 44(6), 1013–1023.
- Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, 136(3), 389. <https://doi.org/10.1037/0096-3445.136.3.389>
- Gori, S., & Facoetti, A. (2015). How the visual aspects can be crucial in reading acquisition? The intriguing case of crowding and developmental dyslexia. *Journal of Vision*, 15(1), 15.11.18. <https://doi.org/10.1167/15.1.8>

- Goswami, U., Fosker, T., Huss, M., Mead, N., & Szucs, D. (2011). Rise time and formant transition duration in the discrimination of speech sounds: The Ba-Wa distinction in developmental dyslexia. *Developmental Science*, 14(1), 34–43. <https://doi.org/10.1111/j.1467-7687.2010.00955.x>
- Hämäläinen, J. A., Salminen, H. K., & Leppänen, P. H. (2013). Basic auditory processing deficits in dyslexia: systematic review of the behavioral and event-related potential/field evidence. *Journal of Learning Disabilities*, 46(5), 413–427. <https://doi.org/10.1177/0022219411436213>
- Hari, R., & Kiesilä, P. (1996). Deficit of temporal auditory processing in dyslexic adults. *Neuroscience Letters*, 205(2), 138–140. [https://doi.org/10.1016/0304-3940\(96\)12393-4](https://doi.org/10.1016/0304-3940(96)12393-4)
- Hari, R., & Renvall, H. (2001). Impaired processing of rapid stimulus sequences in dyslexia. *Trends in Cognitive Sciences*, 5(12), 525–532. [https://doi.org/10.1016/S1364-6613\(00\)01801-5](https://doi.org/10.1016/S1364-6613(00)01801-5)
- Hathaway, Q. A., Roth, S. M., Pinti, M. V., Sprando, D. C., Kunovac, A., Durr, A. J., et al. (2019). Machine-learning to stratify diabetic patients using novel cardiac biomarkers and integrative genomics. *Cardiovascular Diabetology*, 18(1), 1–16. <https://doi.org/10.1186/s12933-019-0879-0>
- Hendren, R. L., Haft, S. L., Black, J. M., White, N. C., & Hoelt, F. (2018). Recognizing psychiatric comorbidity with reading disorders. *Frontiers in Psychiatry*, 9, 101. <https://doi.org/10.3389/fpsy.2018.00101>
- Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, 47(6), 674–689. <https://doi.org/10.1016/j.cortex.2010.07.010>
- Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of Experimental Child Psychology*, 77(1), 30–60. <https://doi.org/10.1006/jecp.1999.2553>
- Kabat, M. H., Kane, R. L., Jefferson, A. L., & DiPino, R. K. (2001). Construct validity of selected automated neuropsychological assessment metrics (ANAM) battery measures. *The Clinical Neuropsychologist*, 15(4), 498–507. <https://doi.org/10.1076/clin.15.4.498.1882>
- Kahta, S., Schiff, R. (2016). Implicit learning deficits among adults with developmental dyslexia. *Annals of Dyslexia* 66, 235–250. <https://doi.org/10.1007/s11881-016-0121-7>
- Knoch, U., & Chapelle, C. A. (2018). Validation of rating processes within an argument-based framework. *Language Testing*, 35(4), 477–499. <https://doi.org/10.1177/0265532217710049>
- Laasonen, M., Service, E., & Virsu, V. (2001). Temporal order and processing acuity of visual, auditory, and tactile perception in developmentally dyslexic young adults. *Cognitive, Affective, & Behavioral Neuroscience*, 1(4), 394–410. <https://doi.org/10.3758/CABN.1.4.394>
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1), 119. <https://doi.org/10.1037/0033-295X.106.1.119>
- Leong, V., & Goswami, U. (2014a). Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing. *Hearing Research*, 308, 141–161. <https://doi.org/10.1016/j.heares.2013.07.015>
- Leong, V., & Goswami, U. (2014b). Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Frontiers in Human Neuroscience*, 8, 96. <https://doi.org/10.3389/fnhum.2014.00096>
- Lervåg, A., & Hulme, C. (2009). Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science*, 20(8), 1040–1048. <https://doi.org/10.1111/j.1467-9280.2009.02405.x>
- Lonergan, A., Doyle, C., Cassidy, C., MacSweeney Mahon, S., Roche, R. A., Boran, L., & Bramham, J. (2019). A meta-analysis of executive functioning in dyslexia with consideration of the impact of comorbid ADHD. *Journal of Cognitive Psychology*, 31(7), 725–749. <https://doi.org/10.1080/20445911.2019.1669609>
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Sage Publications.
- Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., Conceicao, E.L.T. and di Palma, M.A. (2020), “Package robustbase”. <http://robustbase.r-forge.rproject.org/>. Accessed 10 November 2020
- Marinelli, C., Iaia, M., Cassibba, R., Traficante, D., Zoccolotti, P., & Angelelli, P. (2020). La valutazione del linguaggio orale e scritto e del profilo neuropsicologico in bambini bilingui Dati di riferimento per la scuola primaria. *Psicologia Clinica dello Sviluppo*, 1-35. <https://doi.org/10.1449/98294>
- Mascheretti, S., Gori, S., Trezzi, V., Ruffino, M., Facoetti, A., & Marino, C. (2018). Visual motion and rapid auditory processing are solid endophenotypes of developmental dyslexia. *Genes, Brain and Behavior*, 17(1), 70–81. <https://doi.org/10.1111/gbb.12409>
- McDonald, J. F., & Moffitt, R. A. (1980). The uses of Tobit analysis. *The Review of Economics and Statistics*, 62(2), 318–321.

- McGrath, L. M., Peterson, R. L., & Pennington, B. F. (2020). The multiple deficit model: Progress, problems, and prospects. *Scientific Studies of Reading, 24*(1), 7–13. <https://doi.org/10.1080/10888438.2019.1706180>
- Messick, S. (1979). Test validity and the ethics of assessment. *ETS Research Report Series, 1979*(1), i–43.
- Mir, A., & Dhage, S. N. (2018). Diabetes disease prediction using machine learning on big data of healthcare. In 2018 fourth international conference on computing communication control and automation (ICCUBEA) (pp. 1–6). IEEE Access. Pune, India, 16–18 August 2018. <https://doi.org/10.1109/ICCUBEA.2018.8697439>
- Moll, K., Gobel, S. M., Gooch, D., Landerl, K., & Snowling, M. J. (2016). Cognitive risk factors for specific learning disorder: Processing speed, temporal processing, and working memory. *Journal of Learning Disabilities, 49*(3), 272–281. <https://doi.org/10.1177/0022219414547221>
- Moura, O., Pereira, M., Alfaiate, C., Fernandes, E., Fernandes, B., Nogueira, S., et al. (2017). Neurocognitive functioning in children with developmental dyslexia and attention-deficit/hyperactivity disorder: Multiple deficits and diagnostic accuracy. *Journal of Clinical and Experimental Neuropsychology, 39*(3), 296–312. <https://doi.org/10.1080/13803395.2016.1225007>
- Moura, O., Simões, M. R., & Pereira, M. (2014). Executive functioning in children with developmental dyslexia. *The Clinical Neuropsychologist, 28*(sup1), 20–41. <https://doi.org/10.1080/13854046.2014.964326>
- Mullane, J. C., Corkum, P. V., Klein, R. M., McLaughlin, E. N., & Lawrence, M. A. (2011). Alerting, orienting, and executive attention in children with ADHD. *Journal of Attention Disorders, 15*(4), 310–320. <https://doi.org/10.1177/1087054710366384>
- Murineddu, M., Duca, V., & Cornoldi, C. (2006). Difficoltà di apprendimento scolastico degli studenti stranieri. *Difficoltà di apprendimento, 12*(1), 49–70.
- Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society, 18*(6), 275–285. <https://doi.org/10.1002/cem.873>
- Nicolson, R. I., & Fawcett, A. J. (1990). Automaticity: A new framework for dyslexia research? *Cognition, 35*(2), 159–182. [https://doi.org/10.1016/0010-0277\(90\)90013-a](https://doi.org/10.1016/0010-0277(90)90013-a)
- Nicolson, R. I., & Fawcett, A. J. (2011). Dyslexia, dysgraphia, procedural learning and the cerebellum. *Cortex, 47*(1), 117–127. <https://doi.org/10.1016/j.cortex.2009.08.016>
- Nicolson, R., Fawcett, A. J., & Dean, P. (2001a). Dyslexia, development and the cerebellum. *Trends in Neurosciences, 24*(9), 515–516. [https://doi.org/10.1016/s0166-2236\(00\)01923-8](https://doi.org/10.1016/s0166-2236(00)01923-8)
- Nicolson, R. I., Fawcett, A. J., & Dean, P. (2001b). Developmental dyslexia: The cerebellar deficit hypothesis. *Trends in Neurosciences, 24*(9), 508–511. [https://doi.org/10.1016/s0166-2236\(00\)01896-8](https://doi.org/10.1016/s0166-2236(00)01896-8)
- Nicolson, R. I., Fawcett, A. J., Berry, E. L., Jenkins, I. H., Dean, P., & Brooks, D. J. (1999). Association of abnormal cerebellar activation with motor learning difficulties in dyslexic adults. *Lancet, 353*(9165), 1662–1667. [https://doi.org/10.1016/S0140-6736\(98\)09165-X](https://doi.org/10.1016/S0140-6736(98)09165-X)
- Nigro, L., Jiménez-Fernández, G., Simpson, I. C., & Defior, S. (2015). Implicit learning of written regularities and its relation to literacy acquisition in a shallow orthography. *Journal of Psycholinguistic Research, 44*(5), 571–585. <https://doi.org/10.1007/s10936-014-9303-9>
- Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63*, 427–452. <https://doi.org/10.1146/annurev-psych-120710-100431>
- Omar, K. S., Mondal, P., Khan, N. S., Rizvi, M. R. K., & Islam, M. N. (2019). A machine learning approach to predict autism spectrum disorder. In 2019 International conference on electrical, computer and communication engineering (ECCE) (pp. 1–6). IEEE Access. <https://doi.org/10.1109/ECACE.2019.8679454>
- Overy, K. (2003). Dyslexia and music. From timing deficits to musical intervention. *Annals of the New York Academy of Sciences, 999*, 497–505. <https://doi.org/10.1196/annals.1284.060>
- Pagliarini, E., Scoecchia, L., Granocchio, E., Sarti, D., Stucchi, N., & Guasti, M. T. (2020). Timing anticipation in adults and children with developmental dyslexia: Evidence of an inefficient mechanism. *Scientific Reports, 10*(1), 17519. <https://doi.org/10.1038/s41598-020-73435-z>
- Papadopoulos, T. C., Spanoudis, G. C., & Georgiou, G. K. (2016). How is RAN related to reading fluency? A comprehensive examination of the prominent theoretical accounts. *Frontiers in Psychology, 7*, 1217. <https://doi.org/10.3389/fpsyg.2016.01217>
- Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition, 101*(2), 385–413. <https://doi.org/10.1016/j.cognition.2006.04.008>
- Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., Leopold, D. R., Samuelsson, S., Byrne, B., Willcutt, E. G., & Olson, R. K. (2012). Individual prediction of

- dyslexia by single versus multiple deficit models. *Journal of Abnormal Psychology*, 121(1), 212–224. <https://doi.org/10.1037/a0025823>
- Peters, L., Bulthé, J., Daniels, N., Op de Beeck, H., & De Smedt, B. (2018). Dyscalculia and dyslexia: Different behavioral, yet similar brain activity profiles during arithmetic. *NeuroImage: Clinical*, 18, 663–674. <https://doi.org/10.1016/j.nicl.2018.03.003>
- Peters, L., Op de Beeck, H., & De Smedt, B. (2020). Cognitive correlates of dyslexia, dyscalculia and comorbid dyslexia/dyscalculia: Effects of numerical magnitude processing and phonological processing. *Research in Developmental Disabilities*, 107, 103806. <https://doi.org/10.1016/j.ridd.2020.103806>
- Peterson, R. L., Boada, R., McGrath, L. M., Willcutt, E. G., Olson, R. K., & Pennington, B. F. (2017). Cognitive prediction of reading, math, and attention: Shared and unique influences. *Journal of Learning Disabilities*, 50(4), 408–421. <https://doi.org/10.1177/0022219415618500>
- Peterson, R. L., & Pennington, B. F. (2012). Developmental dyslexia. *The Lancet*, 379(9830), 1997–2007. [https://doi.org/10.1016/S0140-6736\(12\)60198-6](https://doi.org/10.1016/S0140-6736(12)60198-6)
- 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>
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25–42. <https://doi.org/10.1146/annurev.ne.13.030190.000325>
- Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences*, 51, 350–365. <https://doi.org/10.1016/j.cageo.2012.08.023>
- R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *Quarterly Journal of Experimental Psychology*, 61(1), 129–141. <https://doi.org/10.1080/17470210701508822>
- 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(Pt 4), 841–865. <https://doi.org/10.1093/brain/awg076>
- Ranganathan, P., Pramesh, C., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Logistic regression. *Perspectives in Clinical Research*, 8(3), 148.
- Rangvid, B. S. (2007). Sources of immigrants' underachievement: Results from PISA—Copenhagen. *Educational Economics*, 15(3), 293–326. <https://doi.org/10.1080/09645290701273558>
- Raven, J. (2003). Raven progressive matrices. In *Handbook of nonverbal assessment* (pp. 223–237). Boston, MA: Springer US.
- Raven, J. (1956). Guide to the coloured progressive matrices (Sets A, Ab, B). London: Lewis.
- Raven, J. C. (1958). Mill hill vocabulary scale: MHV; Raven's Standard Progressive Matrices: SPM. HK Lewis.
- Reid, A. A., Szczerbinski, M., Iskierka-Kasperek, E., & Hansen, P. (2007). Cognitive profiles of adult developmental dyslexics: Theoretical implications. *Dyslexia*, 13(1), 1–24. <https://doi.org/10.1002/dys.321>
- Reiter, A., Tucha, O., & Lange, K. W. (2005). Executive functions in children with dyslexia. *Dyslexia*, 11(2), 116–131. <https://doi.org/10.1002/dys.289>
- Revelle, W. (2014). Package 'psych'. <http://cran.r-project.org/web/packages/psych/psych.pdf>. Accessed 14 Nov 2020
- Richards, T., Grabowski, T., Boord, P., Yagle, K., Askren, M., Mestre, Z., et al. (2015). Contrasting brain patterns of writing-related DTI parameters, fMRI connectivity, and DTI–fMRI connectivity correlations in children with and without dysgraphia or dyslexia. *NeuroImage Clin*, 8, 408–421. <https://doi.org/10.1016/j.nicl.2015.03.018>
- Richardson, U., Thomson, J. M., Scott, S. K., & Goswami, U. (2004). Auditory processing skills and phonological representation in dyslexic children. *Dyslexia*, 10(3), 215–233. <https://doi.org/10.1002/dys.276>
- Ring, J., & Black, J. L. (2018). The multiple deficit model of dyslexia: What does it mean for identification and intervention? *Annals of Dyslexia*, 68(2), 104–125. <https://doi.org/10.1007/s11881-018-0157-y>
- Rostami, M., Farashi, S., Khosrowabadi, R., & Pouretamad, H. (2020). Discrimination of ADHD subtypes using decision tree on behavioral, neuropsychological, and neural markers. *Basic Clin Neurosci*, 11(3), 359–367. <https://doi.org/10.32598/bcn.9.10.115>
- Sartori, G., Job, R., & Tressoldi, P. (2007). DDE-2: Batteria per la Valutazione della Dislessia e della Disortografia Evolutiva–2, Seconda Edizione. Firenze: Giunti O.S.
- Schnepf, S. V. (2004). How different are immigrants? A cross-country and cross-survey analysis of educational achievement. A cross-country and cross-survey analysis of educational achievement. <https://ssrn.com/abstract=621062> or <https://doi.org/10.2139/ssrn.621062>. Accessed 3 Jun 2021
- Shum, D. H., McFarland, K. A., & Bain, J. D. (1990). Construct validity of eight tests of attention: Comparison of normal and closed head injured samples. *The Clinical Neuropsychologist*, 4(2), 151–162. <https://doi.org/10.1080/13854049008401508>

- Šimundić, A. M. (2009). Measures of diagnostic accuracy: Basic definitions. *EJIFCC*, 19(4), 203.
- Smith, P. J., Need, A. C., Cirulli, E. T., Chiba-Falek, O., & Attix, D. K. (2013). A comparison of the Cambridge Automated Neuropsychological Test Battery (CANTAB) with “traditional” neuropsychological testing instruments. *Journal of Clinical and Experimental Neuropsychology*, 35(3), 319–328. <https://doi.org/10.1080/13803395.2013.771618>
- Smith-Spark, J. H., Henry, L. A., Messer, D. J., Edvardsdottir, E., & Zięćik, A. P. (2016). Executive functions in adults with developmental dyslexia. *Research in Developmental Disabilities*, 53–54, 323–341. <https://doi.org/10.1016/j.ridd.2016.03.001>
- Sommer, C., & Gerlich, D. W. (2013). Machine learning in cell biology—teaching computers to recognize phenotypes. *Journal of Cell Science*, 126(24), 5529–5539. <https://doi.org/10.1242/jcs.123604>
- Stein, J. (2018). What is developmental dyslexia? *Brain Sciences*, 8(2), 26. <https://doi.org/10.3390/brainsci8020026>
- Stein, J., & Walsh, V. (1997). To see but not to read; The magnocellular theory of dyslexia. *Trends in Neurosciences*, 20(4), 147–152. [https://doi.org/10.1016/S0166-2236\(96\)01005-3](https://doi.org/10.1016/S0166-2236(96)01005-3)
- Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language*, 9(2), 182–198. [https://doi.org/10.1016/0093-934x\(80\)90139-x](https://doi.org/10.1016/0093-934x(80)90139-x)
- Tallal, P., & Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in Neurosciences*, 29(7), 382–390. <https://doi.org/10.1016/j.tins.2006.06.003>
- Tarca, A. L., Carey, V. J., Chen, X. W., Romero, R., & Drăghici, S. (2007). Machine learning and its applications to biology. *PLoS Computational Biology*, 3(6), e116. <https://doi.org/10.1371/journal.pcbi.0030116>
- Therneau, T., Atkinson, B., Ripley, B., & Ripley, M. B. (2015). Package ‘rpart’. <cran.ma.ic.ac.uk/web/packages/rpart/rpart.pdf>. Accessed on 20 April 2020
- Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal of Physiology, Paris*, 102(1–3), 120–129. <https://doi.org/10.1016/j.jphysparis.2008.03.007>
- Thomson, J. M., Fryer, B., Maltby, J., & Goswami, U. (2006). Auditory and motor rhythm awareness in adults with dyslexia. *Journal of Research in Reading*, 29(3), 334–348.
- Thomson, J. M., Leong, V., & Goswami, U. (2013). Auditory processing interventions and developmental dyslexia: A comparison of phonemic and rhythmic approaches. *Reading and Writing*, 26(2), 139–161.
- Trevethan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, pliabilitys, and pitfalls in research and practice. *Frontiers in Public Health*, 5, 307. <https://doi.org/10.3389/fpubh.2017.00307>
- Van Reybroeck, M., & De Rom, M. (2020). Children with dyslexia show an inhibition domain-specific deficit in reading. *Reading and Writing*, 33, 907–933. <https://doi.org/10.1007/s11145-019-09986-z>
- Varvara, P., Varuzza, C., Sorrentino, A. C., Vicari, S., & Menghini, D. (2014). Executive functions in developmental dyslexia. *Frontiers in Human Neuroscience*, 8, 120. <https://doi.org/10.3389/fnhum.2014.00120>
- 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>
- Viviani, P., & Stucchi, N. (1989). The effect of movement velocity on form perception: Geometric illusions in dynamic displays. *Perception & Psychophysics*, 46(3), 266–274.
- Vogel, S. J., Banks, S. J., Cummings, J. L., & Miller, J. B. (2015). Concordance of the Montreal cognitive assessment with standard neuropsychological measures. *Alzheimers Dement*, 1(3), 289–294. <https://doi.org/10.1016/j.dadm.2015.05.002>
- Wechsler, D. (2003). *Wechsler intelligence scale for children—Fourth Edition (WISC-IV)*. The Psychological Corporation.
- Willcutt, E. G., Sonuga-Barke, E. J., Nigg, J. T., & Sergeant, J. A. (2008). Recent developments in neuropsychological models of childhood psychiatric disorders. In *Biological child psychiatry* (Vol. 24, pp. 195–226). Karger Publishers. <https://doi.org/10.1159/000118526>
- Wolak, M. (2015). ICC: Facilitating estimation of the intraclass correlation coefficient. R Package Version, 2.3.0. <https://cran.r-project.org/web/packages/ICC/ICC.pdf>. Accessed 14 Nov 2020
- Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, 91(3), 415. <https://doi.org/10.1037/0022-0663.91.3.415>
- Yee, T. W. (2008). The VGAM package. *R News*, 8, 28–39.