

# ■ The I-min Screening Test for Reading Problems in College Students: Psychometric Properties of the I-min TIL

Tânia Fernandes<sup>1\*</sup>, Susana Araújo<sup>1,2</sup>, Ana Sucena<sup>3,4</sup>, Alexandra Reis<sup>2</sup> and São Luís Castro<sup>4</sup>

<sup>1</sup>*Faculdade de Psicologia, Universidade de Lisboa, Lisbon, Portugal*

<sup>2</sup>*Departamento de Psicologia e Ciências da Educação, Institute of Biotechnology and Bioengineering, Centre for Molecular and Structural Biomedicine, Universidade do Algarve, Faro, Portugal*

<sup>3</sup>*Escola Superior de Tecnologia da Saúde do Porto, Instituto Politécnico do Porto, Porto, Portugal*

<sup>4</sup>*Faculdade de Psicologia e de Ciências da Educação, Universidade do Porto, Porto, Portugal*

Reading is a central cognitive domain, but little research has been devoted to standardized tests for adults. We, thus, examined the psychometric properties of the I-min version of Teste de Idade de Leitura (Reading Age Test; I-min TIL), the Portuguese version of Lobrot L3 test, in three experiments with college students: typical readers in Experiment 1A and B, dyslexic readers and chronological age controls in Experiment 2. In Experiment 1A, test–retest reliability and convergent validity were evaluated in 185 students. Reliability was  $>.70$ , and phonological decoding underpinned I-min TIL. In Experiment 1B, internal consistency was assessed by presenting two 45-s versions of the test to 19 students, and performance in these versions was significantly associated ( $r = .78$ ). In Experiment 2, construct validity, criterion validity and clinical utility of I-min TIL were investigated. A multiple regression analysis corroborated construct validity; both phonological decoding and listening comprehension were reliable predictors of I-min TIL scores. Logistic regression and receiver operating characteristics analyses revealed the high accuracy of this test in distinguishing dyslexic from typical readers. Therefore, the I-min TIL, which assesses reading comprehension and potential reading difficulties in college students, has the necessary psychometric properties to become a useful screening instrument in neuropsychological assessment and research. Copyright © 2017 John Wiley & Sons, Ltd.

**Keywords:** neuropsychological assessment; screening test; psychometric properties; adults; developmental dyslexia

Reading is a core domain in neuropsychological assessment. In adults, reading abilities account for a significant proportion of variance in other cognitive domains such as expressive vocabulary (Uttl, 2002), executive functions, attention, working memory and verbal fluency (Johnson et al., 2006), with a stronger predictive value

\*Correspondence to: Tânia Fernandes, Faculdade de Psicologia, Alameda da Universidade, Universidade de Lisboa, 1649-013 Lisboa, Portugal. E-mail: taniapgf@gmail.com

The first author is supported by IF 2013 Program of the Portuguese Foundation for Science and Technology, FCT (ref IF/00886/2013/CPI194/CT0002) and by the Center for Research in Psychology at Universidade de Lisboa (CIPUL). This work was also supported by grants from FCT (COMPETE and FEDER programmes) to the Center for Psychology at Universidade do Porto (PEst-C/PSI/UI0050/2011; FCOMP-01-0124-FEDER-022714), and to the projects PTDC/PSI/110734/2009, PEst-OE/EQB/LA0023/2014. We thank Juliana Branco for her assistance in experimental testing.

than schooling (i.e. years of education; Johnson et al., 2006), even in tests with no obvious relation with reading like the clock drawing test (Santana et al., 2013).

Surprisingly, little research has been devoted to standardized reading tests for adults (Lewandowski et al., 2003; Miller-Shaul, 2005; Nergård-Nilssen & Hulme, 2014; Re et al., 2011; Stevens & Price, 1999; with some exceptions, see, e.g. Warmington et al., 2013). Indeed, at least four main reasons argue for the need of such tests with good psychometric properties. First, *developmental dyslexia*, a specific reading disorder that affects over 80% of those identified as learning disabled (Shaywitz et al., 2008; Vellutino et al., 2004), is a chronic disorder expressed by lifelong deficits in *decoding* (or phonological recoding, i.e. orthographic–phonological mapping; cf. Share, 1995) and in other phonological abilities like phonological awareness, verbal memory and verbal processing speed (e.g. Miller-Shaul, 2005; Nergård-Nilssen & Hulme, 2014; Pennington et al., 1990; Ramus et al., 2003; Re et al., 2011; Reid et al., 2007). Second, generalized access to university and anti-discriminatory policies led to an increase in the number of college students, including those with developmental dyslexia (Lewandowski et al., 2003). Third, the relevance of standardized reading tests extends beyond the clinical domain into research. Although college students are not representative of the population of adult dyslexic readers, research has often been conducted with them (e.g. Ramus et al., 2003; with some exceptions, see, e.g. Nergård-Nilssen & Hulme, 2014) because they offer a methodological advantage; dyslexic college students are massively exposed to written material and thus if they continue to present deficits in literacy-related abilities, these cannot be due to reduced exposure to print. Furthermore, although the reading disorder is typically diagnosed during childhood, current reading skills are important as well. Dyslexic adults with deficits only in phonological decoding show differences in brain morphological structure relative to those with deficits in phonological decoding and text comprehension (Welcome et al., 2011). Finally, the lack of standardized tests for adults is also related with the scarcity of cross-linguistic research on the cognitive profile of adult dyslexic readers (for a discussion, see, e.g. Miller-Shaul, 2005; Nergård-Nilssen & Hulme, 2014; Re et al., 2011; Simmons & Singleton, 2000).

In the present study, we aimed to provide a norm-referenced screening test of reading problems for college students, one that allows to quickly differentiate typical readers from those with potential reading difficulties. We examined the psychometric properties of a brief (1 min) reading comprehension test, originally adapted by Sucena and Castro (2008) to European–Portuguese for 8- to 11-year-old children (i.e. from the second to fifth grade), that is, the *Teste de Idade de Leitura* (TIL) from the original French test L3 of Lobrot (1973).

*Teste de Idade de Leitura* is a sentence comprehension test with a paper and pencil format, arranged in two A5 sheets, each with 18 sentences arranged in column (total of 36 sentences). At the end of each sentence, the last word is missing, and five words are presented in parenthesis. The participant is asked to select by circling which of the words correctly completes each sentence within a time limit of 5 min for as many sentences as possible (for more details, see Sucena & Castro, 2008; Vilhena et al., 2016). For example (the correct word is underlined), Item 1: Pega na saca e vai-me comprar...; *hold the bag and go to buy...* (artes, laranjas, lâminas, lavatórios; arts, oranges, blades; sinks); Item 11: Eles trabalham o dia inteiro e à noite...; *they work all day and at night they...* (olham, quebram, penteiam, descartam, descansam; look, break, comb, discard, rest); Item 18: Durante

a noite, espero que tenhas bons...; *during the night, I hope you have good...* (sonhos, olhos, lápis, sorrisos, peixes; *dreams, eyes, pencil, smiles, fishes*); Item 24: Qual é o teu jogo preferido? Ping-pong, bilhar, dominó ou...; *what is your favourite game? Ping-pong, snooker, domino, or...* (camisas, cartas, malas, focas, mãos; *shirts, cards, bags, seals, hands*); Item 35: O marido de uma filha é para a mãe dessa filha o...; *the husband of a daughter is to the mother of that daughter the...* (gigante, agente, genro, gesso, gente; *giant, agent, son-in-law, plaster, people*).

In the present version for adults, the time limit was shortened to 1 min (hence the name 1-min TIL) in order to prevent ceiling effects. Indeed, with the 5 min limit, 11-year-old fifth-grade children present a high performance of, on average, 80% sentences correctly completed (28.5 out of 36,  $SD=5.0$ ; Sucena & Castro, 2008) and college students are at ceiling (Sucena et al., 2014). Thus, dyslexic readers who are college students would likely perform at ceiling as well. Furthermore, none of the students examined by Sucena et al. had a null performance (of 0%) within 1 min. An additional advantage is that a 1-min test can be applied in any context, does not require much time nor financial resources and is suited for routine use in clinical practice (Stevens & Price, 1999). Time is probably less relevant in clinical assessment than in screening, but reading tests are 'often chosen for quickness at the expense of quality, due to the extensive testing in other areas' (Stevens & Price, 1999, p. 69).

Our selection of TIL was based on practical and theoretical criteria. At the practical level, no norm-referenced test is available in Portugal for the adult population.<sup>1</sup> The TIL has good clinical validity in children (e.g. Sucena et al., 2009), providing percentiles, average scores and standard deviation for each school grade between second and fifth-grade based on a sample of 614 children. TIL was recently used to examine the prevalence of dyslexia in Portuguese second to fourth-grade children, which was found to be of 5.4% (Vale et al., 2011). Sucena et al. (2014) have also used TIL to examine the reading performance of college students at a top Portuguese university and found that 18% had a reading performance similar to the best 9-year-old readers in the fourth grade. Whether this reflects a poor reading performance is unclear as no norm-referenced test is yet available for adults.

At the theoretical level, we selected TIL taking into account that no test offers a pure measure of any specific cognitive ability. Indeed, reading comprehension depends on decoding and linguistic comprehension (i.e. the ability to derive sentence and discourse interpretations from lexical and semantic information at the word level; Hulme & Snowling, 2014), in line with the *simple view of reading* (Gough & Tunmer, 1986; Hoover & Gough, 1990). Decoding depends on letter knowledge and phonological skills during the early stages of reading acquisition, and in later stages, it depends on word identification and phonological decoding (the latter usually assessed with non-word reading; Braze et al., 2015; Hulme & Snowling, 2014). Linguistic comprehension is also a multifactorial construct; it comprises oral language skills that are necessary to comprehend words, but expressive vocabulary has been used as proxy in studies with adults (e.g. Braze et al., 2015). Both types of skills (decoding and listening comprehension) can be dissociated and are causally related to reading comprehension throughout development (García & Cain, 2014). Both are involved in any reading comprehension test. Performance will be a product of their weighting, which, in turn, is also affected by test format (Cutting & Scarborough, 2006; Lewandowski et al., 2003;

Nation & Snowling, 1997; for a meta-analysis about the relative importance of decoding to reading comprehension considering assessment characteristics, see García & Cain, 2014). Note that we aimed for a test that would be strongly saturated in phonological decoding for two reasons. First, a phonological deficit is at the core of developmental dyslexia (Vellutino et al., 2004). Second, phonological decoding plays a major role in reading in European–Portuguese, not only in children (e.g. Seymour et al., 2003; Sucena et al., 2009) but also in fluent adult readers (Lima & Castro, 2010). Typically developing European–Portuguese children show a reading accuracy similar to that of children from deeper orthographies (i.e. orthographies with less grapheme–phoneme consistency), but reading is typically faster in Portuguese (Seymour et al., 2003), because it is an orthography of *intermediate* depth (for details, see, e.g. Lima & Castro, 2010; Sucena et al., 2009), that is, more transparent than French or English but deeper than Italian, Spanish or Dutch. Therefore, decoding may not be as problematic for typically developing Portuguese readers as compared with English-speaking readers. Based on the recent meta-analysis of García and Cain (2014), we chose a reading comprehension test with a multiple choice format (see also, e.g. Nation & Snowling, 1997; Wilson & Lesaux, 2001) that involved silent reading and was time-limited because the strongest association with phonological decoding (assessed with nonword reading) was found for tests with this format. We also preferred a time-limited test because most dyslexic adults who are college students have read extensively and may have compensated for their reading difficulty at the expense of fluency (Elbro et al., 1994; Miller-Shaul, 2005; Shaywitz et al., 2008). Thus, in the I-min TIL, inefficient word decoding would hinder the allocation of available cognitive resources to text comprehension (Cutting & Scarborough, 2006). As TIL was originally developed for children, it only requires basic comprehension skills, and hence, its linguistic content would not pose difficulties to adults. Nevertheless, as shown in the examples aforementioned, its content is not child-specific, and hence, is suited for Portuguese adults too.

The present study comprised three experiments. Experiments 1A and 1B composed the normative sub-study. In Experiment 1A, we examined the performance of typically developing readers to gather norms and the corresponding percentile ranks. *Reliability* was assessed with a test–retest procedure, because the time-limited format prevented the adoption of a split-half procedure. To examine *convergent validity*, and because no benchmark reading comprehension test was available, we used a reading aloud test of isolated items. Experiment 1B was a complementary experiment in which TIL's *internal consistency* was examined. Experiment 2 was the clinical sub-study, on which we investigated *construct validity*, *criterion validity* and *clinical utility* of the I-min TIL using a *case-control* design (i.e. the same number of chronological age controls and dyslexic readers; Fischer et al., 2003). Given that TIL requires basic comprehension skills and is supposed to be strongly saturated in phonological decoding, relative to chronological age controls, dyslexic readers were expected to present lower performance because of their decoding deficit.

In Experiment 2, two facets of *ecological validity* were also considered (Spooner & Pachana, 2006), given that neuropsychological tests should reflect cognitive functioning in real-world situations, and psychometric data alone are no longer sufficient for diagnosing developmental dyslexia (e.g. DSM-5; American Psychiatric

Association, A. P. A, 2013). *Verisimilitude* (i.e. the similarity between the cognitive demands of the test and those of everyday situations) was part of the selection criteria: silent reading is an everyday task for college students and text comprehension is the goal of real-world reading. *Veridicality* (i.e. the extent to which test scores are related to other measures predicting real-world performance) was assessed with a self-report measure of reading (Lefly & Pennington, 2000; Tamboer & Vorst, 2015).

## EXPERIMENT 1 (1A AND 1B)

Experiment 1A and B formed the normative sub-study of the present work. In Experiment 1A, we collected the normative 1-min TIL scores from 185 college students; two subsamples were further tested to examine test–retest reliability and convergent validity. The latter was examined with a reading aloud test of isolated items given that no norm-referenced reading comprehension test is available in Portuguese. Furthermore, prior studies have shown that poor text comprehension by dyslexic adults is consequence and not cause of the decoding deficit, and this deficit is also expressed by poor reading *fluency*, that is, the time taken to correctly read aloud an item (e.g. in Danish: Elbro et al., 1994; in Hebrew: Miller-Shaul, 2005; in Norwegian: Nergård-Nilssen & Hulme, 2014; in Italian: Re et al., 2011; in Polish: Reid et al., 2007; and in English: Simmons & Singleton, 2000). We thus examined convergent validity with the Portuguese version of the reading fluency subtest of the *Differential Diagnosis Dyslexia Maastricht Battery*, 3DM (Pacheco et al., 2014; Reis et al., in preparation), given that phonological decoding should be predominant in the 1-min TIL and in a reading fluency test.

Experiment 1B was a complementary experiment conducted with another group of participants to examine internal consistency.<sup>2</sup> A similar procedure as in Experiment 1A was adopted, with the exception that instead of a single session, TIL was presented in two 45-s sessions in which participants performed the test on half of the items: the first column in the first session and the second column in the second session.

## Method

### Participants

All were college students and European–Portuguese native speakers with no history of neurological, psychiatric or developmental disorders (including dyslexia).

In Experiment 1A, 185 students (102 women, 83 men; similar proportion,  $\chi^2(1) = 1.95$ ,  $p = .16$ ), who were on average 22.8 years old ( $SD = 5.25$ ; range: 18–48), from Universidade do Algarve and Universidade do Porto, Portugal, participated voluntarily and completed the 1-min TIL. From these students, further testing involved two subsamples: 22 (7 women, 15 men;  $\chi^2(1) = 2.91$ ,  $p = .09$ ) with an average age of 21.5 years ( $SD = 0.74$ ) who performed the 1-min TIL again; and 50 (29 women, 21 men;  $\chi^2(1) = 1.28$ ,  $p = .26$ ) with an average age of 22.4 years ( $SD = 4.2$ ) who performed the 3DM reading fluency test.

In Experiment 1B, 19 students (15 women, 4 men), from Universidade de Lisboa, Portugal, on average 21.1 years old ( $SD = 2.48$ ; range: 18–29), volunteered to participate.

All experiments followed the ethical guidelines of the Declaration of Helsinki.

#### *Materials and procedure*

In Experiment 1A and B, TIL was collectively administered in a university seminar room. Before the test, participants were presented with four practice items in a different sheet to ensure they understood the task. For these items, there was no time limit, and feedback was provided.

Next, in Experiment 1A, participants were presented with the A4 sheet (organized into two A5 sheets with 18 items each) and asked to perform the multiple choice sentence completion test as fast and accurately as possible until the allotted time had elapsed (of 1 min, not indicated to participants). They were warned that they would not reach the end of the test and this was no cause to worry.

The same instructions were given in Experiment 1B, but each of the 18-item columns of TIL was printed on a separate A4 sheet (one per sheet) and was presented to the participants one at a time, after having been informed that they would perform the test twice. They were asked to perform the test until the allotted time (45 s) had elapsed. After collecting the first sheet, the second was distributed and participants were again asked to perform the test.

Performance in TIL was computed as the total number of sentences correctly completed.

In Experiment 1A, two subgroups performed another testing session. One subgroup ( $n=22$ ) was contacted to participate in a second session ~1 month afterwards, for test–retest examination; another subgroup ( $n=50$ ) was assessed individually in the 3DM test for convergent validity examination. The 3DM test includes three lists (high-frequency words, low-frequency words, non-words; for details, see Pacheco et al., 2014; Reis et al., in preparation) to be read aloud. These were presented in a fixed order, using E-Prime 1.0 (<http://www.pstnet.com/eprime>). Participants were asked to read aloud as fast and accurately as possible the stimuli presented in the computer screen (for 30 s per list). Performance was computed as the number of items correctly read per list.

## **Results and Discussion**

### *Experiment 1A*

College students were able to correctly complete, on average, 15.53 sentences ( $SD=3.09$ ) of the 1-min TIL, with 95% CI [15.09, 15.98]. The median and mode of performance was 15. Thus, in a brief time limit of 1 min, college students who are typical readers were able to select the correct word of almost half of the sentences presented. This result indicates the appropriateness of this time limit to avoid ceiling and floor effects: none of the students had a performance at ceiling (the highest score observed was 25) or at floor (the lowest score was 7). As shown in Figure 1, the distribution of 1-min TIL scores was symmetrical, with skewness close to 0, that is,  $-0.028$  ( $SE=.18$ ), a single peak at the centre of the distribution and kurtosis of 0.63 ( $SE=.35$ ). Yet, it did not follow a normal distribution, *Shapiro–Wilk test*,  $W=.98$ ,  $p=.002$ , precluding the computation of z-scores (Crawford et al., 2009).

There was no difference in performance between women and men, (*Mann–Whitney U test*,  $p=.654$ ), in line with prior studies with college students



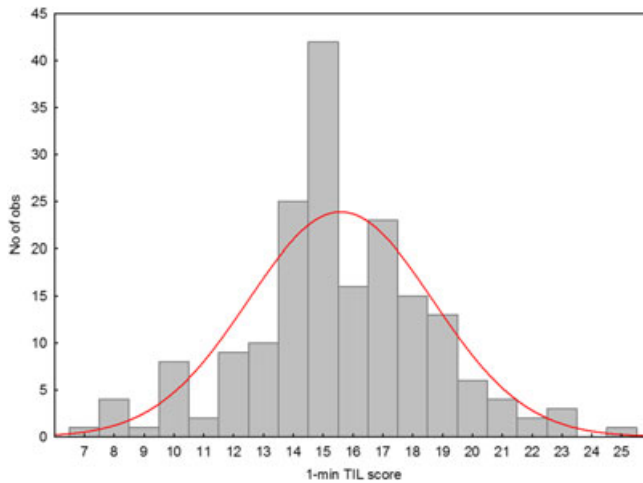


Figure 1. Histogram of the frequency distribution of 1-min TIL scores by college students ( $N = 185$ ). TIL, Teste de Idade de Leitura (Reading Age Test).

(Lewandowski et al., 2003; Uttl, 2002). The scores were not associated with age,  $r_s = .05$ , nor with schooling,  $r_s = -.03$  (both  $ps > .45$ ), probably because participants were fluent readers from a homogenous sample. Indeed, they had on average 15.1 years of education ( $SD = 1.4$ ) and were typical readers with no complaints on reading acquisition or current reading skills. The independence between the 1-min TIL and age and schooling was thus expected.

Percentiles were computed across sex as the percentage of scores falling below the score of interest plus half of those at this score (Crawford et al., 2009). The raw scores and critical percentile ranks are presented in Table 1. Note that test scores correspond to a point estimate of a real-valued score that lies in a specific interval. Therefore, a partial overlap between intervals of close percentile ranks is expected, and percentile values should be interpreted accordingly. For example, as shown in Table 1, there is a partial overlap between the interval estimates of

Table 1. Performance in the 1-min TIL (raw score, i.e. number of correct responses, out of 36) of the normative sample (Experiment 1,  $N = 185$ ): percentile ranks and 95% interval estimates

Raw score	Percentile rank	95% Interval percentile rank	
		Lower	Upper
8	1	0.2	4.8
10	5	2.1	10.4
12	10	6.4	17.1
13	16	10.6	23.0
14	25	16.5	35.8
15	50	30.2	57.7
17	75	61.5	77.9
19	90	80.9	93.6
21	95	91.6	98.2
25	99	97.9	100.0

TIL, Teste de Idade de Leitura (Reading Age Test).

Percentiles 5 and 10, signalling the limits of test precision. Therefore, it was not possible to estimate with sufficient precision the percentile values between these percentile ranks (e.g. Percentile 7). However, for those in Table 1, the 1-min TIL was sufficiently accurate. A 1-min TIL score of 13 was at Percentile 15 with a 95% CI between Percentiles 10.6 and 23.0, and hence, could be an appropriate cut-off score to distinguish typical and dyslexic readers. The rationale of this cut-off score is discussed in Experiment 2, where criterion validity and clinical utility were examined.

For the subgroup that performed the 1-min TIL twice,<sup>3</sup> performance on the second session was significantly higher ( $M=17.91$ ,  $SD=3.42$ ; 95% CI [16.39, 19.43]) than on the first ( $M=15.36$ ,  $SD=2.94$ ; 95% CI [14.06, 16.66]),  $t(21)=2.54$ ,  $p<.001$ , probably because of practice. The correlation between performance in the two sessions was strong,  $r(20)=.71$ ,  $p=.002$ , showing the stability of 1-min TIL scores and demonstrating good test-retest reliability.

For the subgroup that performed the two reading tests, performance in the 3DM reading fluency test<sup>4</sup> was significantly modulated by list (high-frequency words, low-frequency words, non-words: within-participants variable),  $F(2, 98)=249.29$ ,  $p<.001$ ,  $\eta_p^2=.857$ . As expected and already found in Portuguese children (e.g. Moura et al., 2015; Pacheco et al., 2014; Sucena et al., 2009) and adults (Lima & Castro, 2010), college students showed a *frequency* effect, with significantly more high-frequency words read in 30s ( $M=65.90$ ,  $SD=7.74$ ; 95% CI [63.70, 68.10]) than low-frequency words ( $M=58.94$ ,  $SD=9.70$ ; 95% CI [56.18, 61.70]),  $F(1, 49)=78.50$ ,  $p<.001$ , and a *lexicality* effect, with better performance for high-frequency and low-frequency words than for non-words ( $M=43.70$ ,  $SD=8.81$ ; 95% CI [44.80, 49.80]), both  $F_s(1, 49)>300$ ,  $p_s<.001$ . Performance in the 1-min TIL was significantly associated with the three lists of the 3DM test: high-frequency words,  $r=.48$ ; low-frequency words,  $r=.58$ ; and non-words,  $r=.69$ , all  $p_s<.005$ , demonstrating, as expected, that the two tests share reading processes (Elbro et al., 1994; García & Cain, 2014; Lewandowski et al., 2003; Nation & Snowling, 1997). These correlation coefficients are similar to those reported for American college students (Lewandowski et al., 2003), for whom the association between reading comprehension and reading fluency was examined with the Nelson Denny Reading Test, Comprehension and Reading Rate and the Word Attack subtest of the Woodcock Johnson Test of Achievement-Revised. Additionally, a *Principal Component Analysis* (PCA)<sup>5</sup> demonstrated that the 1-min TIL and 3DM tests shared phonological decoding processes, as performance was grounded by a common factor that explained 77.84% of the total variance. Therefore, the correlation matrix analysis and PCA suggest that decoding abilities are a critical component of the 1-min TIL, arguing for the convergent validity of this test.

#### Experiment 1B

College students who performed TIL in 45-s periods were able to correctly complete, on average, 11.95 sentences ( $SD=2.41$ ) of the first column, with 95% CI [10.78, 13.11], and 8.89 sentences ( $SD=2.62$ ) of the second, with 95% CI [7.63, 10.16] (total average performance of 20.84,  $SD=4.75$ , 95% CI [18.55, 23.13]). Importantly, these scores were significantly associated,  $r(17)=.78$ ,  $p<.001$ , demonstrating good internal consistency.



## EXPERIMENT 2

In this clinical experiment, we further investigated construct validity and criterion validity (including clinical utility) of the 1-min TIL. Regarding construct validity, we administered a battery of standardized and experimental measures to dyslexic and chronological age control college students and considered the association of 1-min TIL scores with two sets of measures: (i) with those related with linguistic comprehension and with decoding, because these skills are components of any reading comprehension test, even in adults (García & Cain, 2014; Gough & Tunmer, 1986; Hoover & Gough, 1990; Hulme & Snowling, 2014), and next phonological decoding and listening comprehension were examined as predictors of 1-min TIL scores; and (ii) with measures of phonological abilities (i.e. phonological memory, word reading fluency and *rapid automatized naming*, RAN), visuospatial working memory and non-verbal intelligence. Criterion validity was examined in logistic regression and *receiver operating characteristic* (ROC) analyses on 1-min TIL scores, thus quantifying the clinical utility, specificity and sensitivity of this test. This type of analysis has been successfully applied to determine the classification properties of psychological tests (McFall & Treat, 1999).

We also provided an operational estimate of ecological validity, by examining the association between 1-min TIL scores and the *Adult Reading History Questionnaire*, ARHQ (Lefly & Pennington, 2000; Portuguese version: R. A. Alves & Castro, 2005), given that self-reported measures seem to be strongly correlated with performance in objective reading tests (Nergård-Nilssen & Hulme, 2014; Tamboer & Vorst, 2015).

### Method

#### *Participants*

Thirty-eight European–Portuguese college students ( $M = 23.4$  years old,  $SD = 4.4$ ; range: 18–35) from Universidade do Algarve, Portugal, were paid to participate and gave their informed consent. They had normal-range non-verbal intelligence, with an average non-verbal IQ score of 105.5 ( $SD = 8.8$ ; CI 95% [102.0, 108.0]) in the Portuguese version of Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Wechsler, 1997a; computed over the Block Design, Matrix Reasoning, Picture Completion and Digit Symbol Coding subtests). Nineteen students (12 women, seven men) had a prior formal diagnosis of developmental dyslexia with no other co-morbid disorder, supported by a clinical report. They also presented significant complaints about their current reading skills in the ARHQ (Lefly & Pennington, 2000; Portuguese version: R. A. Alves & Castro, 2005). The 19 typical readers had no known history of neurological, psychiatric or developmental disorders, nor reading complaints, and were matched to the dyslexic group in age,  $F < 1$ , and sex (12 women, 7 men), *Fisher's exact test*,  $p = .63$ .

#### *Materials and procedure*

Participants performed the 1-min TIL and the 3DM reading fluency test as in Experiment 1A. They also performed eight additional tests.

The ARHQ (Lefly & Pennington, 2000) is a self-report measure that examines the risk of reading difficulties in adults based on reading history and current reading habits. The Portuguese version (Alves & Castro, 2005) comprises 25

Likert-type items in a 0–4 scale (the higher the score, the stronger the complaint; maximum score of 100).

The *Vocabulary* subtest of WAIS-III (Wechsler, 1997a) was used to examine expressive lexical knowledge and as an index of general language and listening comprehension skills (Braze et al., 2015). The *Digit Span* subtest (Wechsler, 1997a) examined phonological working memory. Performance in both tests was converted to normalized standard scores.

The *Corsi block* test of the Wechsler Memory Scale (Wechsler, 1997b) examined visuospatial working memory. It is a 3D analogue of the *Digit Span* test; participants are asked to reproduce the sequence of tapping blocks in the same, forward sequence (in the first list), and in backward order (in the second list). Performance was computed as the total number of sequences correctly performed because no norm-referenced scores are available in Portuguese.

The two *phoneme awareness* tasks (Francisco & Faisca, 2012) required phoneme manipulation and have often been used with adults (e.g. Nergård-Nilssen & Hulme, 2014; Ramus et al., 2003; Reid et al., 2007; Vukovic et al., 2004; Wilson & Lesaux, 2001) and children (e.g. Moura et al., 2015). In the *phoneme deletion* task, participants were asked to say the target word (presented aurally) without the initial phoneme in the first block and without the last phoneme in the second block (total of 18 items). Performance was computed as the total of correct responses. In the *spoonerism* task, participants were asked to swap the initial sound of the two aurally presented words and to produce the resulting words (e.g. *no tails* would result in *toe nails*). Performance was computed as the number of words correctly produced (out of 38 words;  $19 \times 2$  words).

Two RAN tasks – with letters and with digits – examined verbal processing speed and lexical access. Participants were required to name aloud the stimuli presented in a  $5 \times 10$  A4 sheet matrix, with no repetition in a row (Alves et al., 2007). Each RAN task comprised five stimuli repeated 10 times each (total of 50 items). Accuracy was at ceiling in both groups; performance was thus computed as the time taken to name all items.

## Results and Discussion

Average performance of the two groups in each test and their comparison are in Table 2, as well as an estimate of the *effect size* for each test (i.e. the ratio of the difference between the average performance of dyslexic and control readers and the *SD* of the control group).

The two groups were adequately matched on non-verbal abilities given that they did not differ in non-verbal IQ nor visuospatial working memory. In contrast, dyslexic readers tended to present worse expressive vocabulary than controls, but the difference was small and is probably a consequence of reading level (Simmons & Singleton, 2000; Vellutino et al., 2004).

In consonance with the results of Experiment 1, the performance of the control group in the I-min TIL fell within the normal range (Table 2) and did not differ from the average result of the normative sample of Experiment 1A (i.e.  $M = 15.53$ ,  $SD = 3.09$ ),  $t < 1$ ,  $p = .38$ . In contrast, dyslexic readers' performance was significantly below that of the normative sample,  $t(17) = -6.12$ ,  $p < .001$ . Nevertheless, for both groups, errors were virtually non-existent and did not

Table 2. Age and cognitive profile of dyslexic and control readers (Experiment 2)

	Dyslexics (n = 19)		Controls (n = 19)		Dyslexics versus controls	Effect size
Age	23.31 (4.60)	23.53 (4.35)	105.05 (8.42)	[100.99–109.11]	$t < 1, p = .89$	–0.05
Non-verbal IQ	105.00 (9.34)	[100.49–109.50]	17.72 (2.42)	[16.52–18.93]	$t < 1, p = .98$	–0.01
Corsi block	17.53 (3.08)	[16.04–19.01]	26.35 (9.84)	[21.29–31.41]	$t = 10.22, p < .001$	–0.08
ARHQ <sup>a</sup>	67.74 (11.86)	[58.02–69.45]	11.26 (2.21)	[10.20–12.33]	$t = -1.90, p = .06$	<b>3.80</b>
Vocabulary <sup>b</sup>	9.89 (2.23)	[8.89–10.97]	11.37 (2.54)	[10.14–12.59]	$t = -2.12, p = .04$	–0.62
Digit span <sup>c</sup>	9.47 (2.95)	[8.05–10.90]	16.42 (2.24)	[15.33–17.50]	$t = -3.90, p < .001$	–0.74
Phoneme deletion <sup>c</sup>	12.84 (3.30)	[11.25–14.43]	31.79 (5.15)	[29.30–34.28]	$t = -3.27, p = .002$	–1.60
Spoonerisms <sup>d</sup>	21.95 (2.77)	[16.53–27.76]	17.32 (2.06)	[16.32–18.30]	$t = 5.98, p < .001$	–1.90
RAN letters <sup>f</sup>	25.05 (5.25)	[22.52–27.58]	16.95 (3.32)	[15.34–18.55]	$t = 6.30, p < .001$	<b>3.76</b>
RAN digits <sup>f</sup>	24.84 (4.34)	[22.75–26.93]	157.53 (17.99)	[148.86–166.20]	$t = -6.66, p < .001$	<b>2.37</b>
3DM reading fluency <sup>e</sup>	104.26 (29.85)	[89.87–118.65]	60.16 (5.49)	[57.51–62.80]	$t = -6.15, p < .001$	–2.96
HF words <sup>g</sup>	42.94 (10.88)	[37.70–48.19]	54.05 (7.73)	[50.33–57.78]	$t = -6.36, p < .001$	–3.13
LF words <sup>g</sup>	34.16 (11.24)	[28.74–39.57]	43.32 (6.86)	[40.01–46.62]	$t = -6.24, p < .001$	–2.57
Nwd <sup>g</sup>	27.16 (8.95)	[22.84–31.47]	15.95 (2.04)	[14.96–16.93]	$t = -5.60, p < .001$	–2.35
I-min TIL	11.16 (3.11)	[9.66–12.81]				–2.35

Note. SD in parentheses. 95% CI in brackets. Significant differences in bold.

HF, high-frequency words; LF, low-frequency words; Nwd, non-words; RAN, rapid automatized naming; TIL, Teste de Idade de Leitura (Reading Age Test); WAIS-III, Wechsler Adult Intelligence Scale-Third Edition; 3DM, Differential Diagnosis Dyslexia Maastricht Battery; SD, standard deviation; CI, confidence interval; ARQH, Adult Reading History Questionnaire.

<sup>a</sup>Maximum score: 100.

<sup>b</sup>WAIS-III normalized standard score ( $M = 10, SD = 3$ ).

<sup>c</sup>Maximum score: 18.

<sup>d</sup>Maximum score: 38.

<sup>e</sup>Summed score across lists.

<sup>f</sup>Time in seconds.

<sup>g</sup>Number of items read correctly in 30 s (maximum: 75).

differentiate dyslexic from typical readers. This pattern of results agrees with the observation of Sucena et al. (2014) that TIL only requires basic comprehension skills. Furthermore, as shown by the effect sizes reported in Table 2, the strongest difference between groups was found for the reading tests, including the I-min TIL, in line with prior evidence in languages varying in orthographic depth (e.g. Elbro et al., 1994; Re et al., 2011; Vukovic et al., 2004; Wilson & Lesaux, 2001). These results indicate the persistence of reading difficulties even in college students massively exposed to print.

The construct validity of the I-min TIL was further assessed by correlation analyses between the I-min TIL and the other cognitive tests. As shown in Table 3, the performance in the I-min TIL was not associated with non-verbal IQ nor visuospatial working memory. In contrast, the strongest correlation was found with the 3DM reading fluency test, across lists:  $r = .83$ ,  $p < .001$ . This result agrees with the correlation pattern found in Experiment 1A, suggesting that the I-min TIL is significantly saturated in decoding abilities.

As shown in Table 3, the I-min TIL was moderately associated with vocabulary. This result is in line with prior studies showing that both skills continue to be reliable predictors of text comprehension in adults (Elbro et al., 1994; García & Cain, 2014; Lewandowski et al., 2003) and with the simple view of reading (Gough & Tunmer, 1986; Hoover & Gough, 1990). The I-min TIL was also moderately associated with phonological working memory (assessed with Digit Span) probably due to the test format itself (García & Cain, 2014; Simmons & Singleton, 2000): participants must hold information in memory to be able to select the correct word that completes each sentence, but both remain visible during response, thus reducing the contribution of phonological memory to I-min TIL performance.

Noteworthy, the association between the I-min TIL and phonological awareness was significant, with correlation coefficients of similar magnitude for the phoneme deletion task,  $r = .38$ ,  $p = .018$ , and the spoonerism task,  $r = .35$ ,  $p = .032$  (cross-task, Table 3). This association is remarkable given that, whereas performance in the I-min TIL involved silent reading and a speeded component, none was involved in the phoneme awareness tasks. Yet, the dyslexic group presented worse phonological awareness than controls (Table 2), and hence, this correlation was possibly inflated by merging the two into a single group. When only typical readers were considered, the association between I-min TIL and phoneme awareness continued to be reliable,  $r(17) = .46$ ,  $p = .045$ .

Rapid automatized naming was more associated with the I-min TIL (across material,  $r = -.77$ ,  $p < .001$ ) than phonological awareness (Table 3). Indeed, apart from the reading tests, the strongest difference between typical and dyslexic Portuguese adults was found for RAN, especially with letters (see also Table 2).

Next, to provide stronger evidence on construct validity, we ran a regression analysis on I-min TIL scores, using phonological decoding and listening comprehension as predictors. Given the moderate-to-strong correlation between performance in several of the cognitive tests examined, namely, those underpinned by phonological decoding (Table 3), to prevent multicollinearity of the predictors and still provide reliable results with such a small sample as the one examined here (e.g. Tabachnick & Fidell, 2007), we first conducted a PCA<sup>7</sup> to extract the latent factor of phonological decoding. It accounted for 73.11% of the total variance of performance in non-word reading, RAN-with-letters, and phoneme awareness

Table 3. Correlation matrix (Pearson coefficients) between performance in the cognitive tests (Experiment 2;  $N = 38$ )

	3DM HF	3DM LF	3DM Nwd	Voc.	PhA	RAN letters (time)	RAN numbers (time)	DS	CB	Non-verbal IQ
I-min TIL	.78***									
3DM HF		.84***	.81***	.54***	.43**	-.74***	-.73***	.43**	.03	.09
3DM LF			.88***	.44*	.47***	-.81***	-.81***	.40*	-.03	.14
3DM Nwd				.51***	.48**	-.83***	-.83***	.55***	.13	.15
Voc.					.57***	-.79***	-.78***	.55***	.09	.14
PhA						-.45**	-.49**	.31 <sup>+</sup>	.18	.26
RAN letters (time)							-.35 <sup>+</sup>	.43*	.02	.22
RAN numbers (time)								-.47**	-.13	-.09
DS									-.44**	-.07
CB										.40*
										.39 <sup>+</sup>

Note. PhA, phonological awareness; DS, digit span; CB, Corsi block; HF, high-frequency; LF, low-frequency; Nwd, non-words; RAN, rapid automatized naming; TIL, Teste de Idade de Leitura (Reading Age Test); 3DM, Differential Diagnosis Dyslexia Maastricht Battery.

\* $p < .05$ .

\*\* $p \leq .01$ .

\*\*\* $p \leq .005$ .

\*\*\*\* $p \leq .001$ .

(the total score in the two tasks). In the regression analysis, the individual standardized scores of phonological decoding derived from the PCA were used as predictor, as well as expressive vocabulary as a proxy of linguistic comprehension<sup>7</sup> (Braze et al., 2015). This model significantly predicted 1-min TIL scores,  $F(2, 34) = 35.04$ ,  $p < .001$ , explaining more than 65% of the variance,  $R^2_{adj} = .66$ . Importantly, both predictors significantly contributed to 1-min TIL scores: phonological decoding,  $\beta_{\text{standardized}} = .69$ ,  $t = 6.30$ ,  $p < .001$ ; listening comprehension:  $\beta_{\text{standardized}} = .22$ ,  $t = 2.05$ ,  $p = .048$ . Therefore, the 1-min TIL is strongly saturated in phonological decoding, but listening comprehension is also a reliable predictor.

To examine the clinical utility of the 1-min TIL, we used logistic regression to determine whether these scores allowed to correctly classify college students with a prior diagnosis of developmental dyslexia versus those without reading difficulties. This was indeed the case,  $X^2_{(1)} = 23.57$ ,  $p < .001$ ,  $NR^2 = .62$ . The 1-min TIL alone was able to correctly classify dyslexic readers in 84.2% of the cases and typical readers in 89.5% of the cases. Overall, 86.8% of the participants were correctly classified on their reading abilities by 1-min TIL scores. These results show the high predictive validity of this screening test.

We also conducted an ROC analysis to examine sensitivity and specificity of the 1-min TIL. Positive and negative predictive values of the test were not calculated given that they depend on the prevalence of the disorder, but the *area under the curve* (AUC), as well as sensitivity and specificity values do not, and hence, can be examined in a case-control design as the present one (e.g. Fischer et al., 2003). The AUC is one of the most useful indexes of the ability of a screening test to discriminate between positive and negative results (McFall & Treat, 1999). Here, it summarizes the discriminative ability of the 1-min TIL: the greater the AUC and closer to 1, the better this test in distinguishing college students with developmental dyslexia from those with no reading deficit. As shown in Figure 2 and in line with the results of the logistic regression, the AUC was .90 ( $SE = .06$ ),  $p < .001$ , 95% CI [.79, 1.0], far above a random classification, which corresponds to a high-level classification accuracy (cf. Hosmer & Lemeshow, 2000). In the majority of cases, 1-min TIL scores distinguished controls from college students with dyslexia, thus demonstrating good criterion validity.

Although the AUC summarizes the discriminability of the test across the full range of scores (Fischer et al., 2003), we also examined the effectiveness of specific cut-off scores (based on the results of the normative experiment; Experiment 1) in distinguishing typical readers from those with a potential reading difficulty. To this aim, we computed the *Youden index* ( $J$ ), which represents the maximal sum of sensitivity and specificity of the test (Youden, 1950), and which, like AUC (Fischer et al., 2003), is independent of the relative size of groups (Youden, 1950). The optimal cut-off of the 1-min TIL (i.e. the maximized Youden  $J$  index: .737) corresponded to a score of 13. With this cut-off score, the 1-min TIL has a specificity of .90 and sensitivity of .84, corroborating the high utility of this test as a screening instrument. In the normative sample (Experiment 1 and Table 1), this score corresponds to Percentile 16 (with 95% CI between Percentiles 10 and 23). When considering more stringent cut-off scores, the sensitivity of 1-min TIL increased to 1 but specificity decreased substantially; for a cut-off score of 10, at Percentile 5 in the normative sample (95% CI between Percentiles 2 and 10), specificity was .42 and the Youden index was .42; for a cutoff score of 12,



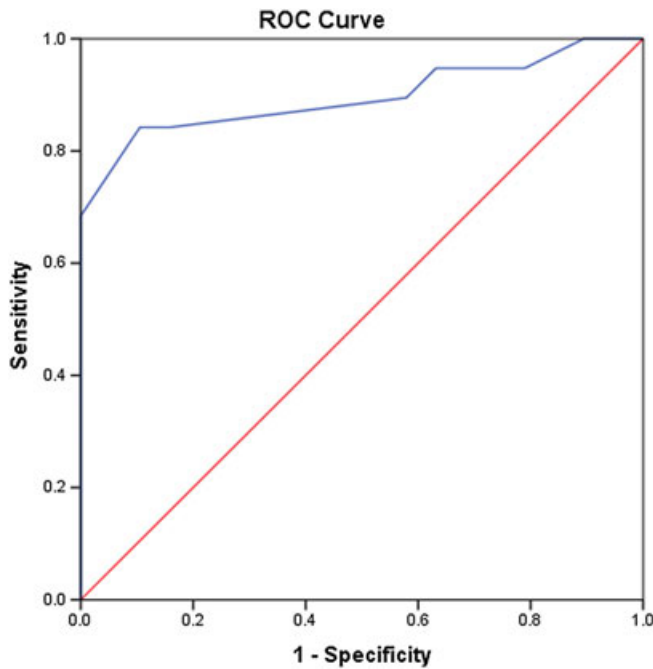


Figure 2. Receiver operating characteristic (ROC) curve of the I-min TIL ( $n = 38$ , cut-off = 13,  $AUC = .896$ ). The red diagonal corresponds to a random classification; the ROC curve (in blue) represents the correct classification by the I-min TIL. Teste de Idade de Leitura (Reading Age Test).

This figure is available in colour online at [wileyonlinelibrary.com/journal/dys](http://wileyonlinelibrary.com/journal/dys)

at Percentile 10 (95% CI between Percentiles 6 and 17), sensitivity was .68 and the Youden index was .68.

Finally, regarding ecological validity, performance in the I-min TIL and the ARHQ was significantly associated,  $r = .734$ ,  $p < .001$ : the lower the I-min TIL scores, the stronger college students' complaints about their reading skills. Indeed, dyslexic participants expressed significantly more and stronger complaints than controls. Self-report measures are able to differentiate dyslexic from typical readers (Nergård-Nilssen & Hulme, 2014; Tamboer & Vorst, 2015). Yet, college students are highly exposed to written material, and hence, they have ample opportunity to become aware of their reading abilities. The association between I-min TIL and AHRQ was inflated given that dyslexic and control readers were merged into a single group. Yet, it was still significant for the dyslexic group alone,  $r(17) = -.46$ ,  $p = .047$ ; for typical readers, the association was negative but no longer reliable,  $r(16) = -.25$ ,  $p = .32$ .

## GENERAL DISCUSSION

The importance of reading assessment in adults is amply recognized (Elbro et al., 1994; Johnson et al., 2006; Lewandowski et al., 2003; Miller-Shaul, 2005; Nergård-Nilssen & Hulme, 2014; Re et al., 2011; Reid et al., 2007; Santana et al., 2013; Shaywitz et al., 2008; Simmons & Singleton, 2000; Stevens & Price, 1999; Warmington et al., 2013; Wilson & Lesaux, 2001). However, no norm-referenced

test was hitherto available for assessing reading comprehension in Portuguese adults. In this study, we investigated the psychometric properties of a brief screening test, that is, the 1-min TIL, which assesses reading comprehension and potential reading difficulties in Portuguese college students. It is a paper and pencil, multiple-choice, sentence completion test, which involves silent reading and is time-limited. We chose this test format because phonological decoding abilities would be preponderant (e.g. García & Cain, 2014; Lewandowski et al., 2003; Nation & Snowling, 1997; Wilson & Lesaux, 2001), which was indeed what we found.

The present results have shown good psychometric properties for this screening test.

Regarding reliability, the 1-min TIL has good test–retest reliability ( $>.70$ ; Experiment 1A) and good internal consistency (.78; Experiment 1B). Convergent validity was examined in Experiment 1 using the 3DM reading fluency test because no criterion reading comprehension test was available and, as expected, both tests were significantly associated and highly saturated in the same latent component, that is, phonological decoding. Additionally, in Experiment 2, the strongest correlation between the 1-min TIL and other cognitive tests was found for the 3DM reading test. Moderate-to-strong associations were also found for phonological memory, phoneme awareness and RAN, which are also underpinned by phonological decoding, albeit to a lesser extent than non-word reading (Araújo et al., 2014; Braze et al., 2015; Hulme & Snowling, 2014; Share, 1995).

Specifically, the 1-min TIL scores were significantly associated with phoneme awareness, showing that the association previously reported in Portuguese children, including dyslexics (e.g. Moura et al., 2015; Sucena et al., 2009), continues to be reliable in young educated adults. The present results also converge with those showing word length effects (a trademark of phonological decoding at grapheme–phoneme unit-level) in Portuguese college students, who flexibly adopt a phonological decoding strategy in specific reading conditions (at least in lexical decision and in reading aloud mixed lists of words and non-words; Lima & Castro, 2010). Note that the association between reading performance and phonemic awareness cannot be explained by considering the latter a proxy of phonological decoding skills. Although the automatic mapping of phonological into orthographic representations facilitates performance in such metalinguistic tasks (Hulme & Snowling, 2014), phonemic awareness is not an orthographic task but one based on the explicit analysis of phonological representations into basic constituents. In contrast, RAN, especially the one with letters, does involve phonological decoding, which explains the stronger association of TIL with RAN than with phoneme awareness found in Experiment 2. In fact, the association between TIL and RAN presumably has several causes: both tests have a speeded component, involve similar eye movements (both require a visual inspection from left to right and up to down) and phonological aspects such as the mapping of a visual (i.e. orthographic in the case of letters, and hence, letter knowledge as well) into a phonological representation and the automatic access to and retrieval of phonological codes (for a recent meta-analysis about RAN and reading, see Araújo et al., 2014; for a review, see Norton & Wolf, 2012). These components also explain, at least partially, the strong association we found between RAN, reading fluency (of non-words and words) and reading comprehension. These results are in line with prior evidence on adults of other

orthographic backgrounds (Miller-Shaul, 2005; Nergård-Nilssen & Hulme, 2014; Reid et al., 2007; Vukovic et al., 2004) and with a recent meta-analysis on child and adolescent readers (<18-year-olds; Araújo et al., 2014).

Furthermore, demonstrating construct validity, the regression analysis of Experiment 2 showed that besides phonological decoding, listening comprehension (with expressive vocabulary as proxy; see, Braze et al., 2015) was a reliable predictor of 1-min TIL scores. This result is consistent with the simple view of reading (Gough & Tunmer, 1986; Hoover & Gough, 1990) and prior evidence (Elbro et al., 1994; Lewandowski et al., 2003; Miller-Shaul, 2005; Nergård-Nilssen & Hulme, 2014; Pennington et al., 1990; Re et al., 2011; Simmons & Singleton, 2000; Vukovic et al., 2004; Wilson & Lesaux, 2001). The strong association between reading fluency and the 1-min TIL is thus explained by automaticity in decoding facilitating reading fluency, which, in turn, frees cognitive resources for use in comprehension (Braze et al., 2015; Cutting & Scarborough, 2006; Vellutino et al., 2004).

Regarding criterion validity and clinical utility, the logistic regression and ROC analyses of Experiment 2 demonstrated the good properties of 1-min TIL. Consistent with prior evidence in languages varying in orthographic depth, Portuguese adult dyslexic readers presented deficits in all reading tests, including the 1-min TIL and in decoding and other phonological tasks (Elbro et al., 1994; Lewandowski et al., 2003; Miller-Shaul, 2005; Pennington et al., 1990; Ramus et al., 2003; Re et al., 2011; Vukovic et al., 2004; Wilson & Lesaux, 2001). In relation to the clinical utility of 1-min TIL, it is important to emphasize that reading, like other cognitive abilities, is measured in a continuum, and thus cut-off scores are relatively arbitrary (e.g. Hulme & Snowling, 2014; Shaywitz et al., 2008), as recognized in international classification manuals like DSM-5 (American Psychiatric Association, A. P. A, 2013). The selection of the cut-off score should thus optimize test sensitivity and specificity (Fischer et al., 2003; Youden, 1950), that is, it should minimize the number of false positives and false negatives. Although DSM-5 suggests that a performance at Percentile 7 in psychometric tests provides the greatest diagnostic certainty, it also indicates that those below the Percentile 16 should be considered, which agrees with the results of Experiment 2. Indeed, a cut-off score of 13 in the 1-min TIL (corresponding to the 15th percentile with a 95% CI between Percentiles 10.6 and 23.0) correctly classified the majority of dyslexic and control college students on their reading skills. Of course, the 1-min TIL does not allow to discriminate in a fine-grained manner the degree of severity of reading difficulties for those with a potential reading deficit. Still, the results of Experiment 2 demonstrate that it can be used as a screening test with high accuracy. Given that the 1-min TIL is highly saturated in phonological decoding and its linguistic content required basic comprehension skills, a low 1-min TIL score will more likely be due to a phonological deficit than to a comprehension deficit. Also, the 1-min TIL is a screening test and not one that by itself could provide a diagnosis of developmental dyslexia.

The original version of the test for children has a 5-min time limit, but the present results demonstrated the adequacy of the 1-min time limit for college students: no typical reader performed at ceiling (the highest score was 25), no dyslexic reader performed at floor (the lowest score was 6) and 1-min TIL scores allowed to discriminate the two groups with high accuracy (AUC = .90; Youden index,  $J = .74$ ). However, not all 36 items of the test were evaluated, because the

highest score did not reach the maximum possible. Yet, in Experiment 1B, internal consistency of the test was .78, demonstrating that performance by the normative sample on the second half of items was highly correlated with performance on the first half. It is unclear whether extending the time limit (e.g. for 2 min) would lead to a significant increase in psychometric quality. Future studies might examine this possibility. Yet, the 1-min time limit is supported by the criterion validity found here.

Regarding ecological validity, in Experiment 2, we found that 1-min TIL scores were associated with subjective reading measures, the ARHQ scores. This result is consistent with prior studies (Elbro et al., 1994; Nergård-Nilssen & Hulme, 2014; Tamboer & Vorst, 2015), but the strong association we found for the dyslexic participants is probably because they had a formal diagnosis of developmental dyslexia and were likely to be aware of the severity of their reading difficulties. The veridicality of the 1-min TIL should thus be further examined, considering other measures than just a questionnaire of subjective reading complaints. We were not able to examine the association between 1-min TIL scores and the neuropsychological assessment at the time of diagnosis because it was made at various ages (from 7 to 8 years old until late adolescence) and based in a heterogeneous set of tests. Furthermore, it was not possible to examine the association between TIL scores and the current reading deficit because no criterion norm-referenced test was available. Future studies should thus examine the association between TIL scores, academic records, clinical assessment, self-reported measures of reading and in adults with the full spectrum of reading skills. Nevertheless, the present results demonstrate the ecological validity of 1-min TIL.

Despite the limitations discussed so far, the present study shows that the 1-min TIL has good psychometric properties and clinical utility. It could be adapted to other languages and orthographies given that it assesses the core cognitive domain affected in developmental dyslexia. Other potential uses could be to monitor the reading abilities of young adults, to test the efficacy of reading interventions in dyslexic readers (who are college students) and to ensure the presence of reading deficits in fundamental research on developmental dyslexia. Furthermore, given that TIL norm-references are available for 8- to 11-year-old Portuguese children (Sucena & Castro, 2008), and the present study provides norms for college students (between 18 and 48 years old), this test might be used to examine reading progress in the long-term. Future studies should thus examine whether TIL would be effective in any of these fronts. In sum, the 1-min TIL is brief and inexpensive, it reflects real-life reading, and, more important, it is a novel, non-redundant instrument, with good psychometric properties that is appropriate to screen for potential reading difficulties in Portuguese college students.

## NOTES

1. In a search conducted in August 2016 in Thomson Reuters, Web of Knowledge, and in NCBI, PubMed with the expressions 'Portuguese adults' and 'reading comprehension test', no results were retrieved. When the expression 'reading test' was used, the only study retrieved was on a Portuguese irregular word reading test aimed at estimating the

- premorbid intelligence level (Alves et al., 2012). It was conducted with 124 16- to 86-year-old adults, but only a small group were college students or had a university 3degree.
2. We thank an anonymous reviewer for this suggestion.
  3. The 1-min TIL scores of this subgroup followed a normal distribution in the first session,  $W = .93$ ,  $p = .12$ , and in the second session,  $W = .93$ ,  $p = .14$ , with homogenous variances, Levene's test:  $F_s < 1$ .
  4. The variance of scores in each list of the 3DM fluency test was homogenous,  $F_s \leq 1$ .
  5. We first ensured the appropriateness of the PCA: the KMO of sampling adequacy was .78, and the Bartlett's test showed a significant correlation between the measures examined,  $X^2(6) = 150.20$ ,  $p < .001$ , justifying data reduction.
  6. We ensured the appropriateness of the PCA: the KMO of sampling adequacy was .59, and the Bartlett's test showed a significant correlation between the measures examined,  $X^2(3) = 47.70$ ,  $p < .001$ , justifying data reduction.
  7. These variables were not collinear ( $VIF = 1.26$ ; tolerance = .79), and the residuals were not auto-correlated (i.e. the Durbin–Watson test,  $D-W = 1.98$ ,  $p > .05$ ).

## REFERENCES

- Alves, L., Simoes, M. R., & Martins, C. (2012). The estimation of premorbid intelligence levels among Portuguese speakers: The Irregular Word Reading Test (TeLPI). *Archives of Clinical Neuropsychology*, 27(1), 58–68. doi:10.1093/arclin/acr103.
- Alves, R. A., Branco, M., Pontes, A., & Castro, S. L. (2007). Avaliação das dificuldades de leitura e escrita: Desenvolvimento da Bateria Fonológica da Universidade do Porto [Assessment of reading and spelling problems: Development of Porto's phonological assessment battery]. *Educação: Temas e Problemas*, 4, 203–222.
- Alves, R. A., & Castro, S. L. (2005). Despistagem da dislexia em adultos através do Questionário História de Leitura [Screening of dyslexia in adults through the adult reading history questionnaire]. *IberPsicologia: Anales de la Revista de Psicología General y Aplicada*, 10(8.9).
- American Psychiatric Association, A. P. A. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5™* (5th ed.). Arlington, VA, US: American Psychiatric Publishing, Inc..
- Araújo, S., Reis, A., Petersson, K. M., & Faisca, L. (2014). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, 107(3), 868–883. doi:10.1037/edu0000006.
- Braze, D., Katz, L., Magnuson, J., Mencl, W. E., Tabor, W., Van Dyke, J., ... Shankweiler, D. (2015). Vocabulary does not complicate the simple view of reading. *Reading and Writing*, 1–17. doi:10.1007/s1145-015-9608-6.
- Crawford, J. R., Garthwaite, P. H., & Slick, D. J. (2009). On percentile norms in neuropsychology: Proposed reporting standards and methods for quantifying the uncertainty over the percentile ranks of test scores. *The Clinical Neuropsychologist*, 23(7), 1173–1195. doi:10.1080/13854040902795018.
- Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10(3), 277–299. doi:10.1207/s1532799xssr1003\_5.
- Elbro, C., Nielsen, I., & Petersen, D. (1994). Dyslexia in adults: Evidence for deficits in non-word reading and in the phonological representation of lexical items. *Annals of Dyslexia*, 44(1), 203–226. doi:10.1007/BF02648162.
- Fischer, J. E., Bachmann, L. M., & Jaeschke, R. (2003). A readers' guide to the interpretation of diagnostic test properties: Clinical example of sepsis. *Intensive Care Medicine*, 29(7), 1043–1051. doi:10.1007/s00134-003-1761-8.
- Francisco, A., & Faisca, L. (2012). Bidirectional association between vocabulary and phonology? A correlational study. Paper presented at the VII meeting of the Portuguese Association of Experimental Psychology, University of Lisbon.

- García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, 84(1), 74–111. doi:10.3102/0034654313499616.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10. doi:10.1177/074193258600700104.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127–160. doi:10.1007/BF00401799.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: John Wiley & Sons.
- Hulme, C., & Snowling, M. J. (2014). The interface between spoken and written language: developmental disorders. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 369(1634). doi:10.1098/rstb.2012.0395.
- Johnson, A. S., Flicker, L. J., & Lichtenberg, P. A. (2006). Reading ability mediates the relationship between education and executive function tasks. *Journal of the International Neuropsychological Society*, 12(01), 64–71. doi:10.1017/S1355617706060073.
- Lefly, D. L., & Pennington, B. F. (2000). Reliability and validity of the adult reading history questionnaire. *Journal of Learning Disabilities*, 33(3), 286–296.
- Lewandowski, L. J., Coddling, R. S., Kleinmann, A. E., & Tucker, K. L. (2003). Assessment of reading rate in postsecondary students. *Journal of Psychoeducational Assessment*, 21(2), 134–144. doi:10.1177/073428290302100202.
- Lima, C. F., & Castro, S. L. (2010). Reading strategies in orthographies of intermediate depth are flexible: Modulation of length effects in Portuguese. *European Journal of Cognitive Psychology*, 22(2), 190–215. doi:10.1080/09541440902750145.
- Lobrot, M. (1973). *Lire avec épreuves pour évaluer la capacité de lecture* (ESF ed.). Paris: ESF.
- McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology*, 50, 215–241. doi:10.1146/annurev.psych.50.1.215.
- Miller-Shaul, S. (2005). The characteristics of young and adult dyslexics readers on reading and reading related cognitive tasks as compared to normal readers. *Dyslexia*, 11(2), 132–151. doi:10.1002/dys.290.
- Moura, O., Moreno, J., Pereira, M., & Simoes, M. R. (2015). Developmental dyslexia and phonological processing in European Portuguese orthography. *Dyslexia*, 21(1), 60–79. doi:10.1002/dys.1489.
- Nation, K., & Snowling, M. J. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, 67(3), 359–370. doi:10.1111/j.2044-8279.1997.tb01250.x.
- Nergård-Nilssen, T., & Hulme, C. (2014). Developmental dyslexia in adults: Behavioural manifestations and cognitive correlates. *Dyslexia*, 20(3), 191–207. doi:10.1002/dys.1477.
- 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. doi:10.1146/annurev-psych-120710-100431.
- Pacheco, A., Reis, A., Araújo, S., Inácio, F., Petersson, K. M., & Faísca, L. (2014). Dyslexia heterogeneity: Cognitive profiling of Portuguese children with dyslexia. *Reading and Writing*, 27, 1529–1545. doi:10.1007/s11145-014-9504-5.
- Pennington, B. F., van Orden, G. C., Smith, S. D., Green, P. A., & Haith, M. M. (1990). Phonological processing skills and deficits in adult dyslexics. *Child Development*, 61(6), 1753–1778. doi:10.1111/j.1467-8624.1990.tb03564.x.
- 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, 841–865. doi:10.1093/Brain/Awg076.
- Re, A. M., Tressoldi, P. E., Cornoldi, C., & Lucangeli, D. (2011). Which tasks best discriminate between dyslexic university students and controls in a transparent language? *Dyslexia*, 17(3), 227–241. doi:10.1002/dys.431.
- Reid, A. A., Szczerbinski, M., Iskierka-Kasperek, E., & Hansen, P. (2007). Cognitive profiles of adult developmental dyslexics: Theoretical implications. *Dyslexia*, 13(1), 1–24.



- Reis, A., Castro, S. L., Inácio, F., Pacheco, A., Araújo, S., Santos, M., & et al., e. (in preparation). Versão Portuguesa da Bateria 3DM para avaliação da leitura e da escrita. [Portugue Version of the 3DM Battery for reading and writing assessment].
- Santana, I., Duro, D., Freitas, S., Alves, L., & Simões, M. R. (2013). The Clock Drawing Test: Portuguese norms, by age and education, for three different scoring systems. *Archives of Clinical Neuropsychology*, 28(4), 375–387. doi:10.1093/arclin/act016.
- Seymour, P. H. K., Aro, M., Erskine, J. M., Wimmer, H., Leybaert, J., Elbro, C., ... network, C. A. A. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2), 143–174. doi:10.1348/000712603321661859.
- Share, D. L. (1995). Phonological recoding and self-teaching – sine-qua-non of reading acquisition. *Cognition*, 55(2), 151–218. doi:10.1016/0010-0277(94)00645-2.
- Shaywitz, S. E., Morris, R., & Shaywitz, B. A. (2008). The education of dyslexic children from childhood to young adulthood. *Annual Review of Psychology*, 59, 451–475. doi:10.1146/annurev.psych.59.103006.093633.
- Simmons, F., & Singleton, C. (2000). The reading comprehension abilities of dyslexic students in higher education. *Dyslexia*, 6(3), 178–192. doi:10.1002/1099-0909(200007/09)6:3<178::AID-DYS171>3.0.CO;2-9.
- Spooner, D. M., & Pachana, N. A. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology*, 21(4), 327–337. doi:10.1016/j.acn.2006.04.004.
- Stevens, K. B., & Price, J. R. (1999). Adult reading assessment: Are we doing the best with what we have? *Applied Neuropsychology*, 6(2), 68–78. doi:10.1207/s15324826an0602\_2.
- Sucena, A., Carneiro, J. F., & de Almeida, F. G. (2014). Assessing reading abilities of mechanical engineering college students: A prospecting study. *International Journal of Engineering Education*, 30(2), 378–387.
- Sucena, A., & Castro, S. L. (2008). *Aprender a ler e Avaliar a Leitura*. [Learning how to read and the assessment of reading]. Coimbra: Almedina.
- Sucena, A., Castro, S. L., & Seymour, P. (2009). Developmental dyslexia in an orthography of intermediate depth: The case of European Portuguese. *Reading and Writing*, 22(7), 791–810. doi:10.1007/s11145-008-9156-4.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon/Pearson Education.
- Tamboer, P., & Vorst, H. C. (2015). A new self-report inventory of dyslexia for students: Criterion and construct validity. *Dyslexia*, 21(1), 1–34. doi:10.1002/dys.1492.
- Uttl, B. (2002). North American adult reading test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology*, 24(8), 1123–1137. doi:10.1076/jcen.24.8.1123.8375.
- Vale, A. P., Sucena, A., & Viana, F. L. (2011). Prevalência da dislexia entre crianças do 1.º Ciclo do Ensino Básico falantes do português europeu [Prevalence of dyslexia in European–Portuguese speaking children of the 1st–4th grade]. *Revista Lusófona de Educação*, 18, 45–56.
- 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. doi:10.1046/j.0021-9630.2003.00305.x.
- Vilhena, D. A., Sucena, A., Castro, S. L., & Pinheiro, A. M. V. (2016). Reading test–sentence comprehension: An adapted version of Lobrot's lecture 3 test for Brazilian Portuguese. *Dyslexia*, 22(1), 47–63. doi:10.1002/dys.1521.
- Vukovic, R. K., Wilson, A. M., & Nash, K. K. (2004). Naming speed deficits in adults with reading disabilities: A test of the double-deficit hypothesis. *Journal of Learning Disabilities*, 37(5), 440–450.
- Warmington, M., Stothard, S. E., & Snowling, M. J. (2013). Assessing dyslexia in higher education: The York Adult Assessment Battery-Revised. *Journal of Research in Special Educational Needs*, 13(1), 48–56. doi:10.1111/j.1471-3802.2012.01264.x.
- Wechsler, D. (1997a). *The Wechsler Adult Intelligence Scale – Third Edition (WAIS-III)*. San Antonio, Tx: The Psychological Corporation.

- Wechsler, D. (1997b). *The Wechsler memory scale-third edition*: San Antonio, TX: The Psychological Corporation.
- Welcome, S. E., Chiarello, C., Thompson, P. M., & Sowell, E. R. (2011). Reading skill is related to individual differences in brain structure in college students. *Human Brain Mapping*, 32(8), 1194–1205. doi:10.1002/hbm.21101.
- Wilson, A. M., & Lesaux, N. K. (2001). Persistence of phonological processing deficits in college students with dyslexia who have age-appropriate reading skills. *Journal of Learning Disabilities*, 34(5), 394–400.
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32–35. doi:10.1002/1097-0142.