

UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA

DOUGLAS DE ARAÚJO VILHENA

Avaliação da habilidade de leitura de crianças do ensino fundamental: medidas direta e
indireta

Belo Horizonte

2015

DOUGLAS DE ARAÚJO VILHENA

**Avaliação da habilidade de leitura de crianças do ensino fundamental:
medidas direta e indireta**

Dissertação apresentada ao Programa de Pós-Graduação em Psicologia da Universidade Federal de Minas Gerais, como parte dos requisitos para obtenção do grau de Mestre em Psicologia.

Área de concentração: Desenvolvimento Humano

Orientadora: Prof^a. Ângela Maria Vieira Pinheiro

Belo Horizonte

2015

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

150

V711a

2015

Vilhena, Douglas de Araújo

Avaliação da habilidade de leitura de crianças do ensino fundamental [manuscrito] : medidas direta e indireta /

Douglas de Araújo Vilhena. - 2015.

88 f.

Orientadora: Ângela Maria Vieira Pinheiro.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Faculdade de Filosofia e Ciências Humanas.

Inclui bibliografia.

1. Psicologia – Teses. 2. Livros e leitura - Teses.
3. Alfabetização - Teses. I. Pinheiro, Ângela Maria Vieira. II. Universidade Federal de Minas Gerais. Faculdade de Filosofia e Ciências Humanas. III. Título.



UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA



FOLHA DE APROVAÇÃO

AVALIAÇÃO DA HABILIDADE DE LEITURA DE CRIANÇAS DO ENSINO FUNDAMENTAL: MEDIDAS DIRETA E INDIRETA

DOUGLAS DE ARAUJO VILHENA

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em PSICOLOGIA, como requisito para obtenção do grau de Mestre em PSICOLOGIA, área de concentração DESENVOLVIMENTO HUMANO, linha de pesquisa Cognição e Linguagem.

Aprovada em 20 de março de 2015, pela banca constituída pelos membros:

Prof(a). Angela Maria Vieira Pinheiro - Orientador
Universidade Federal de Minas Gerais

Prof(a). José Francisco Soares
Universidade Federal de Minas Gerais

Prof(a). Augusto Buchweitz
PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO

Prof(a). Marcos Antônio da Cunha Santos
Universidade Federal de Minas Gerais

Belo Horizonte, 20 de março de 2015.

SUMÁRIO

Resumo Geral	7
Overall Abstract	8
ARTIGO 1: Reading Performance Screening Test: An adapted version of Lobrot’s L3 test for Brazilian Portuguese	9
Abstract	10
Introduction	11
Methods	15
Results	19
Table 1. <i>L3, TIL and TDL – frequencies of: i) distractor type; ii) number of words and letters in the test and iii) position of the target word in the response alternatives.</i>	20
Table 2. <i>Similarity between sentences in the French (L3), the Portuguese (TIL) and the Brazilian (TDL) tests.</i>	21
Discussion	26
Conclusion	27
Conflict of Interest	27
Funding	27
References	27
Appendix 1. <i>The Lecture 3 test of the ORLEC battery: Distractors classified according to their proximity to the sentence, the target word and the other distractors.</i>	31
ARTIGO 2: Validation and Standardization of the Brazilian Reading Performance Screening Test	37
Abstract.....	38
Introduction	39
Methods	41
Table 1. <i>Sample Frequencies According to Age Groups, Gender, and School Grade</i>	42
Results	45
Table 2. <i>Pearson Correlation Between Reading, General Cognitive Ability, Behavior, and Demographic Variables</i>	47
Table 3. <i>TDL’s Percentile Norms for Public State Schools (Raw Scores) According to the Child’s School Grade and Chronological Age</i>	48

<i>Figure 1. Box plot of the five subgroups identified by cluster analysis based on the score of TDL by school grade</i>	49
<i>Table 4. Cross-cultural Comparison With the Normative Data From the L3 test, TIL and TDL</i>	51
Discussion	51
Conclusion	55
Conflict of Interest	55
References	56
ARTIGO 3: EACOL (Scale of Evaluation of Reading Competence by the Teacher): Modification and Validation	62
Abstract	63
1. Introduction	64
2. Method	67
<i>Table 1. Breakdown of the student sample according to age, sex, and grade</i>	68
3. Results	71
<i>Table 2. Pearson correlation between EACOL, reading, general cognitive ability, and behavior</i>	74
<i>Figure 1. Two-step cluster analysis for Reading Aloud (RA) and Silent Reading (SR) items</i>	76
<i>Chart 1. EACOL number of items for each reader category and for the situations Reading Aloud (RA) and Silent Reading (SR)</i>	76
<i>Table 3. Percentile norms and classification for raw EACOL scores by school grade</i> ...	77
4. Discussion	78
5. Conclusion	80
6. Conflict of Interest	81
7. Funding	81
8. References	81
TESTE DE DESEMPENHO DE LEITURA	88
ESCALA DE AVALIAÇÃO DA COMPETÊNCIA DE LEITURA PELO PROFESSOR (EACOL) – FORMA A	90
ESCALA DE AVALIAÇÃO DA COMPETÊNCIA DE LEITURA PELO PROFESSOR (EACOL) – FORMA B	91

RESUMO GERAL

Vilhena, D. A. (2015). Avaliação da habilidade de leitura de crianças do ensino fundamental: medidas direta e indireta. Dissertação de Mestrado, Programa de Pós-Graduação em Psicologia, Universidade Federal de Minas Gerais, Belo Horizonte.

A dissertação é composta por três artigos interligados. O primeiro descreve o processo de adaptação do teste francês *Lecture 3* de Lobrot, renomeado para Teste de Desempenho de Leitura (TDL). Como continuação, o segundo artigo apresenta a validação psicométrica e normatização do TDL. Já o terceiro, relata modificações estruturais realizadas na Escala de Avaliação da Competência em Leitura pelo Professor (EACOL), com subsequente validação e normatização do instrumento. A adaptação do TDL se deu pela demanda de um teste que avaliasse a leitura silenciosa das crianças do Ensino Fundamental e, também, para prover validação concorrente para a sub-escala de leitura silenciosa da EACOL. Foram coletados dados de 484 crianças de 7 a 11 anos (2º ao 5º ano escolar), entre novembro e dezembro de 2013, por meio de sete instrumentos de avaliação de leitura, cognição geral e comportamento social. Tanto o TDL quanto a EACOL demonstraram serem instrumentos robustos na avaliação da competência leitora de crianças típicas e atípicas, o que agiliza o trabalho de pesquisadores e clínicos na detecção de crianças com baixo e alto desempenho escolar.

Palavras-chave: avaliação da leitura, alfabetização, avaliação pelo professor, escala de avaliação da competência em leitura pelo professor, teste de desempenho de leitura

OVERALL ABSTRACT

Vilhena, D. A. (2015). Reading skills assessment of elementary schoolchildren: direct and indirect measures. Dissertação de Mestrado, Programa de Pós-Graduação em Psicologia, Universidade Federal de Minas Gerais, Belo Horizonte.

The dissertation consists of three articles interconected. The first describes the process of adaptation of the French Lobrot's *Lecture 3* test, renamed to Reading Performance Screening Test (Teste de Desempenho de Leitura – TDL). As a continuation, the second article presents the psychometric validation and standardization of TDL. The third, reports structural changes made in the Scale of Evaluation of Reading Competence by the Teacher (Escala de Avaliação da Competência em Leitura pelo Professor – EACOL), with subsequent validation and standardization of instrument. The adaptation of the TDL started due to the demand of a test to assess the silent reading of elementary school children to provide concurrent validation for the silent reading subscale of EACOL. Data were collected from 484 children 7-to-11 years (2nd-to-5th school year) between November and December 2013, by seven instruments to evaluate reading, general cognition and social behavior. Both the TDL and the EACOL showed to be robust instruments in the assessment of reading competence of schoolchildren, helping researchers and clinicians to quick detect students with low and high academic performance.

Keywords: reading assessment, literacy, assessment by the teacher, Escala de Avaliação da Competência em Leitura pelo Professor, Teste de Desempenho de Leitura.

ARTIGO 1

Reading Performance Screening Test: An adapted version of Lobrot's L3 test for Brazilian Portuguese

Douglas de Araújo Vilhena¹, Ana Sucena², São Luís Castro³, Ângela Maria Vieira Pinheiro¹

1. Department of Psychology, Universidade Federal de Minas Gerais, Brazil
2. School of Allied Health Sciences, Instituto Politécnico do Porto, Portugal
3. Faculty of Psychology and Educational Sciences, Universidade do Porto, Portugal

Reading Performance Screening Test: An adapted version of Lobrot's L3 test for Brazilian Portuguese

Abstract

Our aim was to adapt Lobrot's Lecture 3 reading test to a Brazilian cultural-linguistic context. This adapted version is called the Reading Performance Screening Test (Teste de Desempenho de Leitura – TDL) and was developed using the European Portuguese adaptation of L3 as a reference. The present study was conducted in seven steps: 1. classification of the response alternatives of L3 test; 2. adaptation of the original sentences into Brazilian Portuguese; 3. back-translation; 4. adaptation of the distractors from TDL; 5. configuration of TDL; 6. pilot-study; and 7. Validation and standardization. In comparison with L3, TDL included new linguistic and structural variables, such as frequency of occurrence of the distractors, gender neutrality and position of the target words. The instrument can be used for a collective screening or individual clinical administration purposes to evaluate the reading ability of 2nd-to-5th-grade and 7-to-11-years-old students.

Keywords: reading assessment, ORLEC L3 Test, Teste de Idade de Leitura, Teste de Desempenho de Leitura, silent reading.

Introduction

The evaluation of reading ability is fundamental for early intervention in children who struggle with learning at school. Such evaluation is especially relevant in Brazil due to the low scores on national and international scholastic assessments achieved by our population. Only 56% of 8-year-old children are fully literate (Todos pela Educação, 2013), with 11% of people from 15 to 24 years old unable to understand or produce the texts they need despite having attended school (Instituto Paulo Montenegro, 2011). Internationally, for instance according to the Programme for International Student Assessment – PISA (Organisation for Economic Co-operation and Development, 2013), Brazil was among the worst countries for reading (ranking 55 out of 65).

These disappointing results are not due to limited opportunities for attending school, but rather to the ineffective instructional strategies in schools and poor socio-economic background, especially within the public educational system (Duncan & Seymour, 2000; Soares, 2004). Children under such circumstances are at permanent risk of reading failure and need to have their learning screened frequently.

Thus, the construction or adaptation, and subsequent validation, of instruments designed to allow early identification of reading problems has become an important investment. Therefore, the current work aims to adapt the Lecture 3 test of the ORLEC battery, which is an instrument that measures basic reading skills (word recognition and understanding), to the Brazilian context. This tool is widely used in both educational and research contexts of francophone countries and in Portugal.

Lecture 3 test of the ORLEC battery

Originally constructed in French, the ORLEC battery was proposed by Lobrot (1967, 1980) to evaluate the writing (OR – orthographe) and reading (LEC – lecture) efficiency of children from 7 to 13 years old. The reading portion of the battery is divided into four tests as

follows: 1. Lecture 1: reading a short text aloud; 2. Lecture 2: silent reading of isolated words followed by a semantic association judgment; 3. Lecture 3: silent reading of incomplete sentences; and 4. Lecture 4: silent reading of a long text followed by questions.

Lecture 3 (L3) is a reliable screening of students' reading ability and can be administered quickly to groups of pupils or individually. It consists of 40 items of increasing difficulty, formed by a sequence of single incomplete sentences, each followed by a choice of five words as alternative completions. Among these alternatives, only one can correctly complete the sentence. The remaining alternatives are distractor stimuli that share phonological, orthographic or semantic similarity with the target word (Piérart & Grégoire, 2004; Reybroeck & Hupet, 2009).

The child's task is to select the word that is meaningful within the sentence. First, there are four training items that are used to demonstrate how the task should be performed (at this stage, the correct response is explicitly indicated to the child). The other 36 sentences are then completed individually and in silence within a 5-min time limit, without any help from the instructor. The test assesses the relationship between performance and speed because the result corresponds to the number of items correctly answered in 5 minutes.

According to the author of the L3 and of all subsequent studies about it, the test measures both the decoding and semantic components of reading. Contrary to this conception, we argue for the replacement of the term "decoding" with "lexical word recognition" because in order to grasp the meaning of sentences rapidly and efficiently, the reader must have already passed the decoding phase. In this early reading phase, meaning is accessed indirectly via phonological mediation. The child, being engaged in the effortful and time consuming grapheme-phoneme conversion process (decoding) of recognizing words, is left with few resources to direct to the accessing of the meaning of the words. **In addition, L3 requires the choice of a target word among distractors, which competes with decoding to use the working**

memory. Therefore, successful performance in the test requires quick lexical recognition of words.

Although the test is almost fifty years old, the **instrument** is of interest due to its design, ease administration and psychometric properties. As for the latter, in studies with Belgium monolingual French-speaking elementary school children, while Mousty and Leybaert (1999) demonstrated the good sensitivity of the instrument for 2nd and 4th graders (no floor or ceiling effects were encountered), Piérart and Grégoire (2004), with a sample of 2989 children (3rd to 6th graders), provided new norms for L3 and demonstrated its high consistency (Cronbach's $\alpha = .94$, Spearman-Brown split-half coefficient = .98).

The L3 test has been the basis for the **develop** of other tests, such as the *Collective Test of Reading Efficacy* in Spain and the *Reading Age Test* in Portugal, which we will review in the following section for structural comparison purposes.

Collective Test of Reading Efficacy

The Collective Test of Reading Efficacy (Test Colectivo de Eficacia Lectora – TECLE) (Carrillo & Marín, 2009; Marín & Carrillo, 1999) has been used since 1997 for screening purposes to detect Castilian-speaking students with delayed reading. This test is part of the DIS-ESP5 battery (Carrillo & Alegría, 2009; Luque et al., 2012).

Similar to the L3, TECLE is conducted in 5 minutes and evaluates the child's ability to manipulate information that has increasing syntactic, semantic and orthographic complexity. Another similarity between the tests is the type of distractors, which can have phonological, orthographic and semantic functions. Despite these similarities, the TECLE has some important differences from the L3, such as a larger number of incomplete sentences ($N = 64$), fewer alternative choices ($N = 4$), and the presence of at least one pseudoword as a distractor for each item. These differences make the TECLE a completely new test, preventing the structural comparison of its results with those of the L3 or the present work.

Reading Age Test (TIL) – European Portuguese

The European Portuguese adaptation of Lobrot's L3 test, the Reading Age Test (TIL – Teste de Idade de Leitura) (Sucena & Castro, 2010), was undertaken in 2004. At the time, there were no instruments with normative data in Portugal designed to assess reading age or to screen for reading difficulties. The L3 test was chosen for three main reasons: (i) it was a thoroughly tested instrument, widely adopted by both researchers and clinicians in French-speaking countries, (ii) its language shares with Portuguese the same Romance origin, thus allowing for a more straightforward translation/adaptation process and (iii) it assesses reading speed and reading comprehension.

A detailed analysis of the L3 test was conducted with special attention to the types of distractors – visual, phonological, semantic or no proximity to target words. Then, the test was translated and adapted to European Portuguese, maintaining whenever possible the same type of distractor and the same average length (in number of words). As in the original Lobrot test, the sentences were made to have an increasing number of words throughout the test. Finally, the last step consisted of a validation study where the TIL was administered to 614 children and norms for 2nd–5th graders were gathered.

Currently, the TIL is published in Portugal by Almedina. It has been adopted by the scientific and educational communities as an instrument to assess reading age in children from 8 to 11 years old. Recently, initial studies have been conducted to enable the use of the TIL with the adult population, specifically to assess reading skills (Sucena, Carneiro & Almeida, 2014) and to screen for dyslexia in college students (the 1-min TIL; Fernandes et al., 2014). As this instrument has the same structure as the L3, it was also taken as reference in our adapted version.

The present study aims to describe the procedure for the adaptation of the French L3 test to a Brazilian cultural-linguistic context. Furthermore, it will compare the L3 with the versions from Portugal (TIL) and Brazil (TDL).

Methods

The present work took into account the International Test Commission Guidelines for Translating and Adapting Tests (ICT, 2005) and the guidelines proposed by Gudmundsson (2009), as they are comprehensive works in this field and because they focus on the various conditions necessary to increase the likelihood of test equivalence. The following steps were taken in the present adaptation process: 1. classification of the response alternatives of the L3 test; 2. adaptation of the original test (sentences and target words) into Brazilian Portuguese; 3. back-translation (from Brazilian Portuguese to French); 4. adaptation of the distractors for the TDL; 5. configuration of the Brazilian version; 6. pilot-study; and 7. Validation and standardization of the final instrument.

Step 1. Classification of the L3 test

This stage consisted of the analysis of the structure of the L3 to uncover the logical patterns of the sentences and response alternatives, as Lobrot did not explicitly indicate how the selection of the test elements was carried out. Both Piérart and Grégoire (2004) and Sucena and Castro (2010) indicate that the selection of the distractors is based on either visual, phonological or semantic similarities. Both studies suggested that this relationship was due to the proximity/distance of the distractors to the target word and to the sentence. In the present study, we realized that this similarity could also be in relation to the other distractors. The classification of the distractors was carried out by two independent psychologists, proficient in both idioms and knowledgeable about the test content. The two classifications were compared, generating a single consensual version.

For the criteria determining *visual proximity*, it was possible to infer that the alternatives must be an equivalent length in number of letters, be orthographically similar (e.g., the presence of digraphs), and have a minimum of three letters in common, regardless of order. For example, in the first item of the test (see Appendix 1), it is possible to note the visual similarity between the target word (*oranges*) and the distractors (*ordures, ombres, ordres*). All of them have at least four letters in common, are of similar length and have the same ending ("es"), though none of them have strong phonological similarity.

The second distinction is *phonological proximity* that occurs when there is similarity between the phonological units of the words. It can be expressed in the form of alliteration or rhyme; for example, training item 4 illustrates rhyme: *accordeur, chanteur* and *conducteur*. However, we should note that there is also a visual similarity between these words.

The third distinction is *semantic proximity*, which refers to the process of sharing similar semantic frameworks. In training item 4, the meaning of the word *mécanicien* (mechanic) is close in meaning to *véhicule* (vehicle). Another example can be seen in item 24, where the distractors are all names of fish species [carpe (carp), tanche (tench), truite (trout) and perche (perch)].

Finally, there are some distractors with no resemblance to any stimuli. For example, in training item 1, the distractor *loin*, despite being a short word, has no visual, phonological or semantic proximity with the target word *lit* or with the other distractors (*bout, loup, and jour*). It is important to highlight that although the distractor *bout* has no similarity with the target word, it is phonologically close to the distractors "loup" and "jour."

Step 2. Translation into Brazilian Portuguese

Translation of the sentences and target words from French to Brazilian Portuguese was carried by the same professionals as in Step 1, as they were familiar with the culture of both languages. A conceptual translation, rather than the strictly literary one, was emphasized,

taking into account the Brazilian cultural-linguistic context. To reduce discrepancies and for comparison purposes, the Brazilian version also took into consideration the European Portuguese adaptation (TIL). The translated versions were compared, generating a single consensual version.

Step 3. Back-translation

A blind back-translation procedure was performed, where the translators – a Brazilian French teacher (also a psychologist) and a native French speaker, highly proficient in Portuguese – had no access to the L3 and worked independently. The two French versions produced **were compared and to the original** version (L3), and the very few discrepancies were then corrected in the Portuguese text.

Step 4. Adaptation of the distractors

In this step, the distractors (incorrect alternatives) were selected for the 40 items. The alternatives followed the same classification pattern as the original Lobrot test as described in step 1 (see results). The variable “frequency of occurrence of words” according to the Word Frequency Count in Written Brazilian Portuguese (Pineiro, 1996a, 1996b) was included in the selection of the distractors. The purpose of this control is to prevent the activation of a given alternative to guide the response due to its greater familiarity to the reader. In this way, for each grade, the frequency of the target word was classified as high, **medium or low**. For example, in the first sentence of the test for 2nd grade, the target word "laranja (orange)" is a high frequency word, which led to the choice of all distractors being this level of frequency. Another example is the item 14, where the target word médico (doctor) is a medium frequency word for the 3rd grade. Following the same logic, its distractors were selected from the same level of frequency [jacaré (alligator), ninho (nest), senhor (sir), and comércio (market)]. Thus, it is expected, at least with regard to the frequency variable, that all alternatives represent the same level of challenge for children. This classification was not controlled either in the L3 test or in the TIL.

Step 5. Configuration of the Brazilian version

At this stage, as in the L3 and TIL, the items were rearranged according to difficulty level. This classification took into account the length of the sentence and of the response alternatives, the configuration of the distractors, the original position in the test sequence, and the syntactic complexity. **The position of the target words was also controlled.** The distribution of the items by difficulty level for each grade was as follows: 2nd (training until item 9), 3rd (item 10–18), 4th (item 19–27), and 5th (item 28–36); this item classification corresponded to the configuration of the final version of the TDL.

Step 6. Pilot-study

A **pilot study** was conducted to identify flaws and to improve the items. All participants provided informed consent, and the Ethical Committee from the Federal University of Minas Gerais approved the pilot and validation study [Certificate of Appreciation Presentation to Ethics (CAAE): 17754514.6.0000.5149]. The TDL was administered to 5th grade students (n = 43) from a state school in Belo Horizonte, Brazil. Quantitative and qualitative data were analyzed. For the latter, special attention was given to the comments of the students during the test.

Step 7. Validation and Standardization

Vilhena and Pinheiro (2014) meticulously explored this validation step, and provided standardization and a cross-cultural comparison between the TDL and both the Belgium (Piérart & Grégoire, 2004) and Portuguese (Sucena & Castro, 2010) norms. All the procedures and results related to this validation procedure are summarized below.

A sample of 484 students from the 2nd to 5th grade of eight state schools in Belo Horizonte were tested with the following measures: 1) Reading Comprehension subtest (Capellini, Oliveira & Cuetos, 2012), 2) Raven's Coloured Progressive Matrices Test (CPM) (Angelini, Alves, Custodio, Duarte, & Duarte, 1998), 3) Strengths & Difficulties Questionnaire

(SDQ) (Goodman, 1997), 4–5) Word Reading Task (WRT) and the Pseudoword Reading Task (PWRT) (Cogo-Moreira, Ploubidis, De Avila, Mari, & Pinheiro, 2012; Pinheiro, 2013).

Regarding the scoring of the TDL, correct items counted 1 point and incorrect or blank items 0 points. As the TDL evaluates the reading competence as a whole, a dimension reduction by principal component analysis (Carreira-Perpiñán, 1997) was used to incorporate three reading measures (PROLEC and accuracy rate of the WRT and PWRT) to create a robust reading variable, the general reading composite.

Instruction for a collective administration of the TDL

“Dear participants. We are now going to play a game in which you have to complete sentences very quickly. Because of this, please only use a pen or a pencil and not an eraser. Do not turn the sheets I am handing out to you until you are told to do so. Let’s do the first page together. You can see four incomplete sentences each followed by five alternatives. You must select the best word to give meaning to the sentences. Now follow me in silence while I read the first training item aloud (in order not to give away the target word, all alternatives are read with the same intonation). So, which is the best word to complete the sentence? (after the students respond, the correct answer is confirmed). Mark the correct answer with an ‘X’. Now read in silence items 2 to 4 and I will check if you are playing correctly. Now you will have 5 minutes to answer as many items as you can on the second page, like you have just done. During the game, you will not be able to ask any questions. If you do not know an item, just skip it. Now, please start. (allow only 5 minutes of test). Ok, the game is over. Please put down your pen or pencil. Do not worry, if you could not answer all the questions.”

Results

Step 1. Classification of the L3 test

As seen in Table 1, phonological (n = 49) and visual (n = 46) proximity in relation to the target word were the most frequent distractor types in the L3 test. Semantic proximity in

relation to the sentence occurred 37 times, representing 23% of the distractors. In addition, **the test** contains only one occurrence of a homophone (the target word "mer" and the distractor "mère"). Another unique occurrence is the visual proximity to one of the words in the sentence, such as the distractor "chaîne" and the key word of the sentence "chien."

Table 1. *L3, TIL and TDL – frequencies of: i) distractor type; ii) number of words and letters in the test and iii) position of the target word in the response alternatives.*

Category		L3	TIL	TDL
Phonological proximity	to the target	49		49
	to the distractors	9		9
Visual proximity	to the sentence	1		1
	to the target	46		52
	to the distractors	12		12
Semantic proximity	to the sentence	37		32
	to the target	7		7
No proximity	to the target	33		31
	to the distractors	1		1
Homophone	to the target	1		0
Number of words in the test		655	582	596
Number of letters in the test		3598	3118	3285
Target position	A	5	4	8
	B	11	12	8
	C	9	10	8
	D	8	7	8
	E	7	7	8

Step 2. Adaptation of the L3 into Brazilian Portuguese

Table 2 shows the comparisons between the L3 and TDL, between the L3 and TIL and finally between the TDL and TIL. Due to peculiarities of the different languages, any translation of materials from one language to another involves adaptations that can demand

minor to major alterations. As shown, the **semantic** meaning of the majority of the items of the L3 was kept in both the Brazilian and the Portuguese adaptations, with the TDL showing a closer proximity to the L3 than the TIL. For the remaining items, the departure from the ideal of preserving the same meaning as the original version was due to: i) ethical reasons (e.g., items with violent content), ii) the search for precision and ii) the necessity of contextual adjustment. As a result, some of the items of the L3 underwent changes that were slight (only a few words were modified, but the general meaning of the sentence was kept), moderate (the semantic context of the sentence was modified, but its syntactic structure was maintained) or radical (alteration in semantics and syntax).

Table 2. Semantic proximity (by number of sentences) between the French (L3), the Portuguese (TIL) and the Brazilian (TDL) tests.

	L3 with TIL	L3 with TDL	TDL with TIL
Unaltered	22	26	25
Slight change	10	8	9
Moderate change	5	5	4
Radical change	3	1	2

For the comparison between the L3 and TDL, the items in the Brazilian adaptation that underwent a slight change were numbers 2, 10, 12, 16, 22, 23, 24 and 33. In the case of items 2 and 10, for ethical reasons, the negative nature of the sentences was minimized. For example, in item 10 the negative intensity of the sentence "There was a big accident: the train got of the rails" (Il y a eu un grand accident: la locomotive est sortie des rails") was altered to "People got frightened: the train got of the rails".

In the remaining items, item 22 (“Everyone went by car to the forest and then we sat on the grass, where we ate our meals”) illustrates a modification made to make the item more precise (e.g., the word “forest” (forêt) was translated as “park,” which is more precise, as people normally sit in the grass in a park), while item 23 (“They are going to the races on Sunday because they like to see the horses running on the track”) includes a contextual change: “horse running” to “car race,” as horseracing is not part of the reality of Brazilian children.

For the items that required moderate changes in their structure, the alterations were performed for ethical reasons (items 8, 13 and 21) or in an attempt to adapt the items to the Brazilian context (items 13, 29 and 30). Item 13 (“Il est parti à la chasse, c’est pourquoi il a pris son fusil”, [He went out to hunt, and that is why he took his]) exemplifies both situations well, as the use of firearms, apart from being illegal in Brazil, has a violent connotation. Additionally, hunting animals is not a sport in our country. Therefore, this item was changed into “She went out in a hurry, so she forgot her purse.”

Finally, radical changes, were necessary only for item 32. The original sentence was not only unclear, but hard to adapt to Portuguese. The European Portuguese adaptation of this sentence also suffered a radical change that was copied in the Brazilian adaptation.

Another variable that was controlled for in the present adaptation of the L3 was gender. Many sentences in the original version, when translated into Portuguese could be either in the masculine or feminine forms, rather than applicable to both genders. This is the case of the sentence “je suis fatigué” (I am tired), in which “tired” can be, in Portuguese, “cansado” (masculine) or “cansada” (feminine). Items with a determined gender can be ambiguous when presented to the opposite gender and that this ambiguity can lead to a delay in the response. Thus, in the TDL, special care was taken to always use neutral sentences such as “estou com sono” (I’m sleepy) applicable to either gender.

In the comparison between the TDL and TIL, 25 items have the same semantic meaning and 9 items are slightly different, as seen in Table 2. Only 6 items were moderately to radically divergent. As such, the L3 and TIL demonstrate approximately the same difference from the Brazilian adaptation. In other words, the three versions are comparable.

Step 3. Adaptation of the distractors

As seen in Table 1, there were only a few differences between the L3 and TDL. The alternatives that differed from the original version were due to the inability to find a matching word in the Brazilian *Word Frequency Count* list. For example, in item 4, the target word *mer* (sea) and the distractor *mère* (mother) are homophones. Due to the lack of a Portuguese homophone for “mar” (sea), the chosen translation for the distractor was “par” (pair), which has phonological and visual proximity.

Considering the variable frequency of occurrence of words classification, the Brazilian adaptation has 6 items (15%) with high frequency, 6 (15%) with medium frequency and 28 (70%) with low frequency.

Step 4. Back-Translation

First, only the items that maintained the same semantic meaning in the *Translation into the Brazilian Portuguese* step were compared (n = 26). After the semantic comparison of the two back-translations with the original French test, it was found that all the items had the same original meaning, thus they did not require any adjustments. Later, the items that suffered a slight change were compared (n = 8). When one takes into account how few words were modified in the adaptation of these items, the back-translation of each of them corresponded well to the original item.

The TDL has 32 target words (80%) with the same meaning as the original L3 test. This identity of meaning between the versions was confirmed in the back-translation. This means that only 8 target words (20%) had to be modified due to the change in the composition of the

sentence or to adapt to the Brazilian sociocultural context [e.g., jonquilles (daffodils) was translated to “roses”].

Step 5. Configuration of the Brazilian version

The original structure of the test was preserved: 4 training items on the front page, 36 test items on the back page, each item occupying up to two lines, and items arranged according to a gradual increase in difficulty level.

As already mentioned, one of the criteria for the difficulty ranking was the length of each item (number of words and letters in the sentence plus the alternatives). As seen in Table 1, the TDL has fewer words (-9%) and letters (-9%) (the equivalent to 3.6 items) than L3, but more words (2%) and letters (5%) (the equivalent to 2 items) than TIL. The comparison between TIL and L3 shows that the Portuguese test has fewer words (-11%) and letters (-13%) than the French one (equivalent to 4.8 items). Thus, considering that the quantity of information (measured by the item length) matters in time measured instruments, the L3 is the hardest test, followed by the TDL, with the TIL being the easiest.

Table 1 shows that there was a large discrepancy between the targets' positions in the original test and in the TDL, with the target words being twice as likely to be in the second position (B) than in the first position (A). This discrepancy is even larger in the TIL, where the target word is three times more frequently in position B than in position A. In the Brazilian version, the position of all target words are equally distributed (20% occurrence in each position). Moreover, because the alternatives set to high and medium frequencies of occurrence are more easily answered than the low frequency of occurrence alternatives due to the familiarity effect, they were not allocated to positions A or B.

Step 6. Pilot-study

The scores of the TDL (Male: range = 20 to 34, M = 27, SD = 4.6; Female: range = 24 to 36, M = 31, SD = 4.3) were in agreement with Piérart and Grégoire's (2004) norms, as well

as with Sucena and Castro's (2010). The very high Cronbach's Alpha (.92) demonstrated very good reliability of the items. The qualitative analyses conducted consisted of the identification and correction of flaws and dubious distractors.

Step 7. Validation and standardization

The TDL showed a very good internal validity, demonstrated by a schooling effect, $F(3, 480) = 76.7$, $MSE = 56.8$, $p < .001$, (2nd < 3rd < 4th < 5th grade); age effect, $F(4, 479) = 56.7$, $MSE = 57.1$, $p < .001$, (7 < 8 < 9 < 10 < 11 years-old); and by the high Cronbach's alpha (.967). A floor effect occurred in the 2nd grade, as 26% of the students scored fewer than 5 items. The ceiling effect was found only in the 5th grade; 22% of students scored more than 30 items correctly and 6 students (1.3%) scored the maximum number of points.

The good concurrent validation was demonstrated by the moderate-to-strong correlation with all of the reading measures. A strong correlation was found with the TDL and the accuracy rate (total number of correct stimuli read per minute) of Word and Pseudoword Reading Tasks ($r = .840$ and $.787$, respectively) and with the general reading composite ($r = .837$). A moderate correlation was found with the accuracy measures (total number of correctly read stimuli). The moderate correlation with the Reading Comprehension subtest ($r = .582$), demonstrated that comprehension is a variable embedded in TDL. Additionally, a moderate correlation between reading performance and general cognitive ability ($r = .502$) and a mild correlation with the psychiatric behaviors ($r = -.344$) was found. The TDL had the highest correlation with school grade and age when compared to all the instruments used.

The TDL demonstrated adequate ability, attested to by the internal validity, to evaluate global reading performance when classified by school grade (2nd to 5th grade) and chronological age (7–11-years-old). The standardization study demonstrated that the large number of items in the test (36) enabled the clear differentiation and classification of the reading performance of students. A two-step cluster analysis, confirmed by a univariate analysis of variance

(ANOVA) with a Bonferroni correction, suggested the existence of five proficiency groups (reading disability, low performance, average performance, above average performance, high performance). Compared with the Brazilian sample, the Belgium scores were lower in the 3rd grade, but equivalent in the 4th and 5th grades. The norms from Portugal were statistically higher ($p < .01$) than the Brazilian in all the school years compared (2nd–5th grade). For further details concerning this validation and standardization process, see Vilhena and Pinheiro (2014).

Discussion

The objective of this study was to describe the construction procedure of the *Reading Performance Screening Test* (TDL) developed by adapting Lobrot's L3 reading test to a Brazilian cultural-linguistic context. The L3 was chosen because it is an important francophone instrument to evaluate the reading ability of young students. Because of its high consistency, good reliability and updated norms, it has been the basis of developing equivalent tests both in Spain and in Portugal.

The TDL, as with its predecessors, is a decision-making test that measures the reading accuracy (word recognition), speed, vocabulary knowledge and comprehension of written materials. In comparison with the L3, the TDL included new linguistic and structural variables, such as frequency of occurrence of the distractors, gender neutrality and position of the target word.

The instrument has been submitted to a validation and standardization process and the results of that study, reported in Vilhena and Pinheiro (2014), demonstrated a robust schooling and age effects and significant correlations with all tests used to measure reading and general cognitive ability. Additionally, it is a **reliable** measure to evaluate a child's academic year (2nd to 5th grade) and **chronological** reading age (7 to 11 years). The data reveals that the instrument is reliable to access the reading ability of students ranging from weak to high global performance up to the 4th grade. The ceiling effect found in the 5th grade shows that the TDL

presents limitations in discriminating the reading performance of students at advanced levels of schooling.

Conclusion

Due to the rigorous procedure of its adaptation to the Brazilian context and the extra control of variables introduced as well as the good results of the validation study, the TDL has proven to be a reliable instrument for evaluating the global reading competence of students from 2nd–5th grade and 7–11 years old. Because the TDL has many linguistics components embedded in its structure, it allows for the screening of different cognitive functions in a single assessment. Additionally, this instrument, as with its predecessors, can be used for a collective screening or for individual clinical administration purposes to evaluate a child's reading grade. The appendix of this paper is an important tool to use for future adaptations of the L3 to other languages as well as for the construction of new tests.

Conflict of Interest

The authors declare no conflicts of interest.

Funding

This work was supported by the Pró-Reitoria de Pesquisa da Universidade Federal de Minas Gerais and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant No. 134357/2013-2). These bodies had no involvement in the study design; collection, analysis and interpretation of data; writing of the report; or the decision to submit the article for publication.

References

- Carrillo, M.S., & Alegría, J. (2009). Exploración de habilidades fonológicas en escolares disléxicos: teoría y práctica. *Revista de Logopedia, Foniatría y Audiología*, 29, 105–120.
[http://dx.doi.org/10.1016/S0214-4603\(09\)70149-4](http://dx.doi.org/10.1016/S0214-4603(09)70149-4)

- Carrillo, M.S., & Marín, J. (2009). Test de Eficiencia Lectora – TECLE. In A. Cuadro, D. Costa, D. Trias & P. Ponce de León (Eds.), *Evaluación del nivel lector: manual técnico del test de eficacia lectora (TECLE)*. Uruguay: Prensa Médica Latinoamericana. ISBN: 978-9974-5684-2-6
- Duncan, L. G., & Seymour, P. H. K. (2000). Socio-economic differences in foundation level literacy. *British Journal of Psychology*, *91*, 145–166.
- Fernandes, T., Araújo, S., Sucena, A., Reis, A., & Castro, S.L. (2014). Reading skills in young adults: psychometric properties of a screening test of reading in Portuguese college students. Manuscript submitted for publication.
- Génard, N., Mousty, P., Content, A., Alegria, J., Leybaert, J. & Morais, J. (1998). Methods to establish subtypes of developmental dyslexia. In P. Reitsma & L. Verhoeven (Eds.), *Problems and interventions in literacy development*, 163–176. Netherlands: Kluwer Academic Press. http://dx.doi.org/10.1007/978-94-017-2772-3_10
- Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, *61*(2), 29–45.
- International Test Commission (2005). International Test Commission Guidelines for Translating and Adapting Tests. Final version. Retrieved May 20, 2013, from <http://www.intestcom.org/upload/sitefiles/40.pdf>
- Instituto Paulo Montenegro (2011). INAF Brasil 2011, Indicador de Alfabetismo Funcional. Principais resultados. Retrieved 17 November 2014 <http://www.ipm.org.br>
- Lobrot, M. (1967). *Batterie pour mesurer la lecture et l'orthographe, ORLEC*. Beaumont/Oise: Bureau d'études et de recherches.
- Lobrot, M. (1980). *Lire avec e ´preuves pour e ´valuer la capacite ´ de lecture (D-OR-LEC)*. Paris: Editions ESF.

- Luque, J. L., Carrillo, M., Alegría, J., Bordoy, S. & López-Zamora, M. (2012). Ventajas del diagnóstico etiológico de la dislexia evolutiva: Informe automatizado a partir de la batería DIS-ESP. In Navarro, J, Fernández, M^a.T^a, Soto, F.J. y Tortosa F. (Eds.), *Respuestas flexibles en contextos educativos diversos*. Murcia: Consejería de Educación, Formación y Empleo.
- Marín, J. & Carrillo, M. S. (1999). *Test Colectivo de Eficacia Lectora (TECLE)*. Published Manuscript. Departamento de Psicología Básica y Metodología. Universidad de Murcia. Test available at: www.pediatrasandalucia.org/Pdfs/TestEficienciaLectora.pdf
- Mousty, P et Leybaert, J. (1999) Evaluation des habilités de lecture et d'orthographe au moyen de BELEC: données longitudinales auprès d'enfants francophones testés en 2^{ème} et 4^{ème} années. *Revue Européenne de Psychologie Appliquée*, 49, 325–342. <http://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/73073/Details>
- Organisation for Economic Co-operation and Development (2013). PISA –Programme for International Student Assessment. Results in Focus: What 15-year-olds know and what they can do with what they know. Retrieved 12 November 2014 <http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf>
- Piérart, B. & Grégoire, J. (2004). Déchiffrer et coprendre: le test de closure en lecture revisité étalonnage belge du L3 de Lobrot. *Le Langage et l'Homme*, 39 (2), 87–100.
- Pinheiro, A. M. V. (1996a). *Contagem de Frequência de Ocorrência de Palavras Expostas a crianças na faixa pré-escolar e séries iniciais do 1º grau*. São Paulo: Associação Brasileira de Dislexia.
- Pinheiro, A. M. V. (1996b). Word Frequency Count in Written Brazilian Portuguese. In S. Contento (Eds.). *Psycholinguistics as Multidisciplinarily Connected Science*. Cesena: Società Editrice Il Ponte Vecchio, 2, 47–52.

- Reybroeck, M. & Hupet, M. (2009). Acquisition of number agreement: effects of processing demands. *Journal of Writing Research*, 1 (2), 153–172.
- Soares, J. F. (2004). O efeito da escola no desempenho cognitivo de seus alunos. *REICE – Revista Eletrônica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 2(2), 83–104.
- Sucena, A. & Castro, S.L. (2010). *Aprender a Ler e Avaliar a Leitura. O TIL: Teste de Idade de Leitura*. 2nd Edition. Coimbra: Almedina. ISBN: 978-972-40-3919-0
- Sucena, A., Carneiro, J. F. & Almeida, F. G. (2014). Assessing reading abilities of mechanical engineering college students: a prospective study. *International Journal of Engineering Education*, 30 (2), 378–387.
- Todos pela Educação (2013). Instituto Paulo Montenegro/IBOPE, Fund. Cesgranrio, Inep. Retrieved 17 November 2014 <http://www.todospelaeducacao.org.br>
- Vilhena, D. A., & Pinheiro, A. M. V. (2014). Validation and Standardization of the Brazilian Reading Performance Screening Test. Manuscript submitted for publication.

Appendix 1. *The Lecture 3 test of the ORLEC battery: Distractors classified according to their proximity to the sentence, the target word and the other distractors.*

1 Training	Je ferai la vaisselle demain matin, car je suis fatigué et je préfère aller au [I'll do the dishes tomorrow because I'm tired and I'd rather go to]				
	bout [end]	loup [wolf]	jour [day]	lit [bed]	loin [far]
	nT + pD	nT + pD	nT + pD	Target	nT + nD
2 Training	Si on fait marcher trop fort sa radio, on risque de déranger les [If we turn the radio too loud, we risk disturbing the]				
	poissons [fishes]	mains [hands]	coins [corners]	voisins [neighbors]	trains [trains]
	vT	vT	vT	Target	vT
3 Training	Mon frère a fait un voyage en Afrique et a rapporté une très originale [My brother made a trip to Africa and brought a very original]				
	ville [city]	statue [statue]	chaleur [heat]	estrade [platform]	saison [season]
	nT + sS	Target	nT + sS	vT + sS	nT + sS
4 Training	Un homme qui conduit un véhicule s'appelle un [A man who drives a vehicle is called a]				
	mécanicien [mechanic]	compagnon [companion]	accordeur [tuner]	conducteur [driver]	chanteur [singer]
	sT + sS	sT	pT	Target	pT + vT
1 Test	Prends le panier et va m'acheter des [Take the basket and go buy me some]				
	armoires [wardrobes]	oranges [oranges]	ordures [garbage]	ombres [shadows]	ordres [orders]
	nT	Target	vT	vT	vT
2 Test	Si vous mangez ce gâteau, dit ma mère, vous verrez comme il est [If you eat this cake, my mother said, you will see that it is]				
	long [long]	rond [round]	bon [good]	doux [sweet]	chou [cabbage]
	pT	pT	Target	nT + sS	nT + pD
3 Test	Tous les chiens ont quatre [All dogs have four]				
	bouches [mouths]	pattes [paws]	pincés [pliers]	prunes [plums]	oreilles [ears]
	sS	Target	vT	vT	sS

4 Test	J'aimerais aller sur la plage pour me baigner dans la [I would like to go to the beach for a swim in the]				
	guerre [war]	mer [sea]	mère [mother]	marche [march]	marque [brand]
	pT	Target	Homophone	pD	pD
5 Test	La petite fille a mis sa [The little girl put on her]				
	roche [rock]	cloche [bell]	roue [wheel]	rue [street]	robe [dress]
	vT	pD	vT	vD	Target
6 Test	La gare se trouve au milieu de la [The station is in the middle of the]				
	filles [daughters]	ville [city]	bille [marble]	boule [boule]	poule [hen]
	pT	Target	pT	pD	pD
7 Test	Il a ouvert la radio et a écouté les [He turned on the radio and listened to the]				
	nouvelles [news]	chandelles [candles]	voiles [sails]	vitres [windows]	navires [ships]
	Target	pT	nT	vD	nT
8 Test	Il a déchiré son tablier et il s'est fait [He tore his apron and he got]				
	rouler [rolled]	grandir [grewed]	sonder [sounded]	craindre [afraid]	gronder [scolded]
	nT	vT	pT	sT	Target
9 Test	Un endroit où on range les livres s'appelle une [A place where we keep books is called a]				
	pêche [peach]	cuisine [kitchen]	galerie [gallery]	bibliothèque [library]	porte [door]
	nT	sS	sS	Target	nT
10 Test	Il y a eu un grand accident: la locomotive est sortie des [There was a big accident: the locomotive got off the]				
	tiroirs [drawers]	rails [rails]	rayons [rays]	routes [roads]	rangs [ranks]
	nT	Target	vT	sT	vT
11 Test	Ils travaillent toute la journée et le soir ils se [They work all day and at night they]				
	noient [drown]	brisent [break]	sèchent [dry]	répondent [meet]	reposent [rest]
	vT	vT	vT	vT	Target

12 Test	Vous pourriez enlever la poussière avec un [You could remove the dust with a]				
	palais [palace]	balai [broom]	bœuf [beef]	lard [bacon]	four [furnace]
	pT	Target	nT	nT	pT
13 Test	Il est parti à la chasse, c'est pourquoi il a pris son [He went out to hunt, and that is why he took his]				
	outil [tool]	feu [fire]	fusil [shotgun]	gentil [gentil]	foin [hay]
	sS	nT	Target	vT	nT
14 Test	Mon oncle, après de longues études, est devenu [My uncle, after a long time studying, became a]				
	nouille [noodle]	médecin [doctor]	moisi [moldy]	monsieur [sir]	moyen [means]
	nT	Target	nT + vD	sS + vD	vT
15 Test	Il s'est penché sur le puits et il est tombé au [He leaned over the well and fell to the]				
	fond [bottom]	front [front]	frein [brake]	fard [rouge]	four [furnace]
	Target	pT	vT	vT	vT
16 Test	Il fait chaud sur la terrasse, pourquoi ne mettez-vous pas le ...? [It's hot on the terrace, why don't you put up the]				
	paravent [folding screen]	radiateur [heater]	parasol [parasol]	passage [passage]	patin [roller skate]
	pT + sT	sS + nT	Target	vT	vT
17 Test	Quand on est dans la rue, il faut faire très attention aux autos afin de ne pas se faire [When you are in the street, you have to be very careful with cars so you do not get]				
	laver [washed]	transporter [transported]	casser [broken]	pousser [pushed]	écraser [run over]
	nT + pT	sS + pT	sT + pT	pT	Target
18 Test	Quand vous dormirez, j'espère que vous ferez de jolis [When you sleep, I hope you will have sweet]				
	rêves [dreams]	yeux [eyes]	trous [holes]	rires [laughters]	cous [necks]
	Target	nT	nT + vD	sS	nT + vD

19 Test	Parmi tous les jeux préférez-vous le ping-pong, le billard, les dominos, ou les ...? [Of all the games you prefer table-tennis, billiards, dominoes, or]				
	douches [showers]	astres [stars]	bras [arms]	cartes [cards]	cadres [frames]
	vT	vT	nT	Target	pT
20 Test	Il s'est pris la main dans la porte et il s'est mis à pleurer en poussant des [He trapped his hand in the door and he started to cry and]				
	bruits [sounds]	lits [beds]	nuits [nights]	cris [scream]	cas [cases]
	pT + sS	pT	vD	Target	vT
21 Test	Un camarade l'a poussé et il est tombé sur les [A friend pushed him over and he fell on his]				
	roues [wheels]	mains [hands]	nains [dwarfs]	vins [wines]	ponts [bridges]
	nT	Target	pT	pT	sS
22 Test	Tout le monde est parti en voiture jusqu'à la forêt et là, nous nous sommes assis sur l'herbe, où nous avons mangé notre [Everyone went by car to the forest and there we sat on the grass, where we ate our]				
	rat [rat]	rang [rang]	repas [food]	quart [quarter]	pas [no]
	vT	vT	Target	nT	pT
23 Test	Ils comptent aller aux courses dimanche prochain car ils aiment voir les chevaux courir sur la [They are going to the races next Sunday because they like to see the horses run on the]				
	piste [track]	liste [list]	voûte [arch]	route [road]	mine [mine]
	Target	pT	vT	vT + sS	nT
24 Test	Il est arrivé une drôle d'aventure à un pêcheur; il a attrapé une [A funny adventure happened to a fisherman; he caught a]				
	carpe [carp]	tanche [tench]	godasse [boot]	truite [trout]	perche [perch]
	sS	sS	Target	sS	sS
25 Test	Du cratère du volcan s'échappent peu à peu des flots de [The volcano crater gradually released flows of]				
	vague [wave]	lave [lava]	bave [drool]	cave [cave]	rage [rage]
	vT	Target	pT	pT	vT

26 Test	Pourquoi ne vous servez-vous pas d'un couteau pour manger votre ...? [Why don't you use a knife to eat your]				
	vin [wine]	voiture [car]	viande [meat]	voisin [neighbor]	ville [city]
	vT	vD	Target	vD	vT
27 Test	Tous les gens sont sortis de leur maison et ont regardé les dégâts produits par l' [Everyone left their homes and watched the damage caused by the]				
	explosion [explosion]	exposition [exhibition]	ascension [ascension]	expédition [expedition]	exagération [exaggeration]
	Target	pT	pT	pT	pT
28 Test	Nos voisins ont acheté un gros chien méchant qui doit rester devant la porte pour monter la [Our neighbors have bought a big mean dog that stays outside the door to stand]				
	corde [rope]	fuite [escape]	chaîne [chain]	grade [grade]	garde [guard]
	vT	nT	vS	pT	Target
29 Test	C'est l'hiver, et cette nuit sont tombés de gros [It's winter, and that night fell large]				
	flacons [bottles]	cocons [cocoons]	flocons [flakes]	sapins [firs]	sabots [shoes]
	pT	pT	Target	sS	sS
30 Test	Nous sommes allés nous promener dans la forêt et nous avons rapporté des [We went for a walk in the forest and we collected some]				
	chalets [cottages]	champignons [mushrooms]	châtaigniers [chestnut trees]	châteaux [castles]	chapeaux [hats]
	pT	Target	Target + vT	pT + vD	pT + vD
31 Test	C'est le printemps, les bois sont fleuris de [It's spring, the woods are flowered with]				
	quilles [bowling]	jongleurs [jugglers]	jonques [junks]	jonquilles [daffodils]	feuilles [leaves]
	pT	pT + vT	pT	Target	sT + vT + sS
32 Test	La fatigue, le surmenage, ont rendu cette personne [The fatigue, the overwork, made this person]				
	alerte [alert]	petite [small]	aimable [friendly]	maligne [malignant]	souffrante [suffer]
	sS	sS	sS	sS	Target

33 Test	Le prestidigitateur, en plantant un couteau dans la paume de sa main, nous a [The magician, by sticking a knife into the palm of his hand, made us]				
	payés [paid]	effacés [cleared]	fouillés [searched]	effrayés [scared]	ensanglantés [bloody]
	pT	pT	pT	Target	sS + pT
34 Test	Les hommes aiment ce qui est nouveau parce que cela satisfait leur [Men love what is new because it satisfies their]				
	bonté [goodness]	amitié [friendship]	curiosité [curiosity]	vanité [vanity]	justice [justice]
	sS + pT	sS + pT	Target	sS + pT	sS
35 Test	Le mari d'une fille est pour la mère de cette fille un [The husband of a daughter is to the mother of that girl a]				
	géant [giant]	agent [agent]	gendre [son in law]	geôlier [jailer]	gendarme [policeman]
	pT	pT	Target	vT	pT
36 Test	Les réfrigérateurs empêchent la nourriture de [Refrigerators prevent food from]				
	mourir [dying]	rouiller [rusting]	se souiller [being defiled]	geler [freezing]	pourrir [rotting]
	sS + vT	sS + vT	vT	sS	Target

Target: correct alternative; pT: phonological proximity to the Target; pD: phonological proximity to the Distractors; vS: visual proximity to the Sentence; vT: visual proximity to the Target; vD: visual proximity to the Distractors; sS: semantic proximity to the Sentence; sT: semantic proximity to the Target; nT: no proximity to the Target; nD: no proximity to the Distractors.

ARTIGO 2

Validation and Standardization of the Brazilian Reading Performance Screening Test

Douglas de Araújo Vilhena* and Ângela Maria Vieira Pinheiro*

*Universidade Federal de Minas Gerais, Brazil

Abstract

We aimed to validate and standardize Reading Performance Screening Test (Teste de Desempenho de Leitura [TDL]). Students (N = 484) from 2nd to 5th grade of eight state schools in Belo Horizonte participated in the study. T-tests, correlations, analyses of variance, and a cluster analysis were used to evaluate the data. The correlations between the TDL and a general reading composite score were high ($r = .837$), as were the accuracy rate of word ($r = .840$) and pseudoword ($r = .787$) reading tasks. The TDL demonstrated adequate ability to evaluate global reading performance when classified by school grade and chronological age (7–11-years-old). The test is a quick and easy screening for children with delayed reading and higher reading performance.

Keywords: reading diagnosis, reading tests, cross-cultural studies, Teste de Idade de Leitura, ORLEC L3.

Introduction

The Reading Performance Screening Test [Teste de Desempenho de Leitura (TDL)] is an adaptation of Lobrot's Lecture 3 reading test (L3, Lobrot, 1967) to the Brazilian cultural-linguistic context (Vilhena, Sucena, Castro, & Pinheiro, 2014). It evaluates lexical word recognition and semantic components of reading in students from 2nd through 5th grade.

Previous Versions

The L3 test evaluates the silent reading ability of French-speaking children and is part of the writing and reading ORLEC [orthographe (OR) and lecture (LEC)] battery proposed by Lobrot (1967, 1980). It consists of 36 incomplete sentences, followed by a choice of five words for completing the sentence. Only one of the five is the correct answer (target word). The remaining are incorrect alternatives (distractors), and relate to the target word through visual, phonological, or semantic proximity or distance. The sentences are presented in an order of increasing difficulty (number of letters and syntactic complexity).

Since its creation, the ORLEC has been validated and updated norms are available (Génard et al., 1998; Mousty & Leybaert, 1999; Piérart & Grégoire, 2004). Mousty and Leybaert (1999) evaluated 217 monolingual French-speaking children in the 2nd and 4th school year in Belgium. The L3 test demonstrated good sensitivity for these grades because no floor effect was observed in the second year (only 10% did not complete more than 5 items correctly) nor ceiling effect in the fourth year (only 10% of children completed more than 30 items correctly). Later, Piérart and Grégoire (2004) tested 2989 French-speaking Belgian elementary school children (3rd to 6th grade), provided new norms for the L3, and demonstrated its high consistency and good reliability. Additionally, as gender differences in scores in the 3rd and in the 5th grades were found, specific standardized and percentile norms for boys and girls were generated.

The L3 test is often used to evaluate reading ability (e.g., Rousselle & Noel, 2006), dyslexia (e.g., Serniclaes, Heghe, Mousty, Carré, & Sprenger-Charolles, 2004), deaf individuals (e.g., Alegría, Domínguez, & Straten, 2009; Colin, Leybaert, Ecalle, & Magnan, 2013; Leybaert, 2000), or as an exclusion criteria (e.g., Mussolin, Mejias, & Noel, 2010; Reybroeck & Hupet, 2009). Due to its versatility, the L3 has been the basis for the construction of other tests, such as the Collective Test of Reading Efficacy [Test Colectivo de Eficacia Lectora (TECLE)] in Spain and the Reading Age Test [Teste de Idade de Leitura (TIL)] in Portugal.

The TECLE (Carrillo & Marín, 2009; Marín & Carrillo, 1999) has been used since 1997 to screen for delayed reading in Castilian-speaking students. Although the TECLE has many similarities to the L3, it possesses a larger number of items, fewer alternatives, and at least one pseudoword as a distractor for each item.

The TIL is closer to the L3 than the TECLE. It has the same number of items as the original French version and a similar structure. It was administered to 614 children and norms were generated for 2nd through 5th grade children (Sucena & Castro, 2010). As it is one of the only standardized tests in Portugal designed to assess reading age or to screen for reading difficulties in children from 8–11 years old, it has been widely adopted by both the scientific and educational community.

Adaptation of the L3 Test to the Brazilian Portuguese

The L3 test was adapted to Brazilian Portuguese was by Vilhena et al. (2014) using the following steps. First, the sentences and the target words were translated from French to Portuguese in consideration of the Brazilian cultural-linguistic context. Second, the distractors (incorrect alternatives) of the L3 Test were classified by their visual, phonological, or semantic proximity or distance to the target word, sentence, and other distractors; this classification was necessary because no detailed information was available in the published materials of the L3.

Additionally, to prevent a given alternative guiding the response due to its greater familiarity, the selection of the Brazilian Portuguese distractors took into account the variable “frequency of occurrence of words” using the Word Frequency Count in Written Brazilian Portuguese (Pineiro, 1996). Third, a blind reverse translation procedure, in which the translator had no access to the original version of the L3, revealed that all the equivalent items had the same original meaning. Finally, the items of the TDL were rearranged according to a gradual increase in their level of difficulty, which ensured the layout of this final version was the same as the layout of the original L3 test.

For comparison purposes, this process also took into consideration the TIL Test (the European Portuguese adaptation of the L3). The proximity between this version and the Brazilian was maintained as much as possible.

The level of proximity between the TIL, TDL, and L3 was maintained by keeping the meaning of the sentences in the L3 in both the Portuguese (N = 22) and the Brazilian (N = 26) adaptations. The alteration of meaning in the remaining items on the TDL varied from minor to major due to ethical reasons, a search for precision, or contextual adjustment. Finally, the L3 Test is the largest with 3598 letters, followed by the TDL with 3285 letters, and the TIL with 3118.

The present study aimed to validate and standardize the TDL. We also compared the Brazilian norms with those of Belgium and Portugal.

Methods

Participants

Students (N = 484) from the 2nd to the 5th year of eight state schools in Belo Horizonte participated in the study (Table 1). The selection of all schools and participants was random, preventing a biased sample. In each of the 82 classrooms, only six students completed a cognitive test battery. The teachers (N = 81) completed a behavior scale for each student. All

participants provided informed consent, and the Ethical Committee from the Federal University of Minas Gerais approved the study (CAAE: 17754514.6.0000.5149).

Table 1. *Sample Frequencies According to Age Groups, Gender, and School Grade*

Age in years	Gender		School Grade				Total
	Male	Female	2 nd	3 rd	4 th	5 th	
7	25	19	44	0	0	0	44
8	54	66	62	57	1	0	120
9	67	53	0	47	70	3	120
10	65	78	0	3	59	81	143
11	27	30	0	0	0	57	57
Total	238	246	106	107	130	141	484

Instruments

The TDL was administered in groups no larger than 10 children. The test instructions were explained using four sample items. The students were asked to answer the test as fast as they could, in a maximum of 5 minutes, without assistance.

The Word Reading Task (WRT) and the Pseudoword Reading Task (PWRT) are both reading aloud instruments, each consisting of 88 items, and individually administered (Cogo-Moreira, Ploubidis, De Avila, Mari, & Pinheiro, 2012; Pinheiro, 2013). The words vary by a) the frequency of occurrence (high and low frequency words), b) bidirectional regularity (regular and irregular words according to grapheme–phoneme/phoneme–grapheme correspondence), and c) length (short, medium, and long words). The pseudowords were constructed with the same orthographic structure and length of stimuli used in the word task.

On both instruments, two measures were used: accuracy, which is the total number of correctly read words or pseudowords, and accuracy rate, which is the total number of correct words or pseudowords read per minute.

The Reading Comprehension subtest is part of the PROLEC (Provas de Avaliação dos Processos de Leitura) (Capellini, Oliveira & Cuetos, 2012). It consists of four short texts, which investigate students' ability to extract meaning and integrate it with prior knowledge. Each text has four questions (literal and inferential), resulting in 16 questions.

Raven's Coloured Progressive Matrices Test (CPM) (Angelini, Alves, Custodio, Duarte, & Duarte, 1998) was used to measure general cognitive ability through the evaluation of analogic reasoning, which is the ability to infer relations between objects or elements (Pasquali, Wechsler, & Bensusan, 2002). Individual application of the CPM was conducted with 2nd year students and the collective form was used for students from grades 3 to 5.

Psychiatric disorders were assessed by the Strengths & Difficulties Questionnaire (SDQ) (Goodman, 1997). This can be completed by parents or teachers and is a brief behavioral screening for 4- to 16-year-olds (Cury & Golfeto, 2003; Saur & Loureiro, 2012). The single-sided Brazilian version, without the impact supplement, was used (Goodman, 2005). This instrument has 25 items divided in 5 scales: prosocial behavior (empathy/positive relations), emotional symptoms (anxiety/mood), conduct problems (aggression/delinquency), hyperactivity/inattention, and peer relationship problems (withdrawn/social problems).

Statistical Analysis

All analyses were performed using the software IBM SPSS Statistics version 21.0 (IBM, Chicago, Illinois). No outliers were detected using the outlier labeling rule (Tukey, 1977) with a g value of 2.2 (Hoaglin & Iglewicz, 1987).

As TDL evaluates the reading competence as a whole, a dimension reduction by principal component analysis (Carreira-Perpiñán, 1997) was used to incorporate three reading

measures (PROLEC and accuracy rate of the WRT and PWRT) to create a robust reading variable, the general reading composite.

For the standardization study, the norms for TDL were split into school grades and chronological age. Although age and schooling are highly correlated in a child's life, reading ability is primarily dependent on formal schooling and environmental factors in the school context. In addition, data of the correlation analysis demonstrated that the age variable was also a reliable parameter, probably because it is an indicator of neural maturation (Primi, Couto, Almeida, Guisande, & Miguel, 2012).

A cut-off score for the lower reading performance was established as the 25th percentile, based on the study by Génard et al. (1998), which demonstrated that 69 out of 75 dyslexic children scored in the lowest quartile on the L3. Thus, the 25th percentile is a good predictor of reading disability, especially for research purposes. However, it is important to note that Rousselle and Noël (2006) assert that the choice of the percentile 15 as a more conservative criteria score, not only guarantees the diagnosis of reading disability, but also avoids false positives when used for clinical purposes. The Diagnostic and Statistical Manual-5 (DSM-5) (American Psychological Association, 2013) recommends a rigorous cut-off score in the 7th percentile for a Specific Learning Disorder (in this case, with the specifier for impairment in reading). However, when considering an academic skill well below average age, this manual also endorses a more lenient threshold of up to the 25th percentile.

A hierarchical two-step cluster analysis was carried out to verify the number of statistically distinct latent groups in the sample. This method assumes that the distance between two clusters is equivalent to the decrease in log-likelihood function as a result of merging. The Bayesian information criterion (BIC) was established to compare the amount of latent classes, in which small values correspond to better fit.

For the cross-cultural comparison of norms between instruments, the creation of a general group sample (*g*) was used for Piérart and Grégoire (2004) and Sucena and Castro (2010) because these studies provided the mean score and standard deviation for males and females separately (Spiegel, Schiller, & Srinivasan, 2000).

The mean and standard deviation of the general group were calculated from the corresponding values of the male and the female groups and evaluated by t-tests. A pooled SD was used, which assumes that the standard deviations of the two samples are similar.

For comparison purposes, a cut-off point for sample size and study antiquity was established. Only the studies with more than 85 participants were considered. This criteria excluded Mussolin, Mejias, and Noel's (2010) work ($N = 15$, Age = 10,3, $M = 26.6$, $SD = 2.1$) as well as Leybaert's (2000) ($N = 30$, Age = 8.8, $M = 21.7$, $SD = 8.2$). Additionally, the studies published over 10 years ago were excluded, thus eliminating Mousty and Leybaert (1999) (2nd grade: $N = 270$, Age = 7.8, $M = 12.3$, $SD = 5.9$; 4th grade: $N = 270$; Age = 9.8; $M = 22.9$; $SD = 5.7$). With the exception of the present work and Mousty and Leybaert (1999), none of the standardization studies noted if the sample was random or not. Due to the distinct features of the TECLE, it was not possible to compare its standardization with those of L3 or with the Portuguese adaptations. Statistical significance was set at $p < .05$

Results

Validation Study

For internal validity, a univariate analysis, corrected for familywise error with Bonferroni, revealed a significant schooling effect, $F(3, 480) = 76.7$, $MSE = 56.8$, $p < .001$, (2nd < 3rd < 4th < 5th grade), and age effect, $F(4, 479) = 56.7$, $MSE = 57.1$, $p < .001$, (7 < 8 < 9

< 10 < 11 years-old). Additionally, Cronbach's alpha for the TDL was .967, which demonstrates strong internal consistency.

To check for the data distribution, skewness and kurtosis values were divided by the respective standard error, using a criteria of significance of higher than 1.96 (Cramer & Howitt, 2004). When all the school grades were analyzed together, the TDL demonstrated a symmetric (.45) and a platykurtic (3.78) distribution. When the data was split by the school grade, only the 2nd (2.58) and the 5th (2.55) grades showed a significant, although small, skewness. These were confirmed using the Shapiro-Wilk normality test. They were due to the floor (fewer than 5 items correct) and ceiling effects (more than 30 items correct) according to the Mousty and Leybaert's (1999) criterion. Unlike the Mousty and Leybaert study, 26% of the 2nd year students scored fewer than 5 items, which was evidence of a floor effect, **suggesting a high risk for learning disorder in the present sample**. However, similar results were found for the 4th grade, **where** 10% of students chose more than 30 items correctly. The ceiling effect was found only in the 5th grade; 22% of students scored more than 30 items correctly and 6 students (1.3%) scored the maximum number of points.

For a concurrent validation of the TDL, a bivariate correlation showed a significant *p*-value of .001 with all of the reading and the general cognitive ability tests (Table 2). A strong correlation was found with the TDL and the accuracy rate of Word and Pseudoword Reading Tasks ($r = .840$ and $.787$, respectively) and with the general reading composite ($r = .837$). A moderate correlation was found with the accuracy measures, which are untimed measures of the Reading Words Test ($r = .550$) and Pseudowords Reading Test ($r = .570$). There was also a significant moderate correlation with the PROLEC Reading Comprehension Test ($r = .582$), demonstrating that comprehension was embedded in TDL. Additionally, a moderate correlation between reading performance and general cognitive ability ($r = .502$) and a mild correlation with the psychiatric behaviors ($r = -.344$) was found.

As seen in Table 2, the TDL had the highest correlation with school grade and age; therefore, it is a good measure to evaluate children by academic year and reading age. The TDL had a larger correlation with grade and age than the CPM, which has a strong age correspondence due to neural maturation.

Table 2. *Pearson Correlation Between Reading, General Cognitive Ability, Behavior, and Demographic Variables*

		TDL	Grade	Age
	TDL		.566**	.565**
Reading	WRT (accuracy measure)	.550**	.330**	.319**
	WRT (accuracy rate)	.840**	.557**	.570**
	PWRT (accuracy measure)	.570**	.306**	.297**
	PWRT (accuracy rate)	.787**	.509**	.526**
	Text comprehension	.582**	.384**	.380**
	General reading composite	.837*	.528*	.546*
Cognition	CPM	.502**	.445**	.432**
Behavior	Prosocial Behavior	.161*		
	Emotional Symptoms	-.290*		
	Conduct Problems	-.216*		
	Hyperactivity/Inattention	-.375*		
	Peer Relationship Problems	-.126*		
	Total negative behaviors	-.344*		

Note. * $p < .01$, ** $p < .001$. Correlations absent in the table were not significant.

TDL: Reading Performance Screening Test; WRT: Word Reading Task; PWRT: Pseudoword Reading Task; Text Comprehension: PROLEC Text Comprehension subtest; CPM: Coloured Progressive Matrices scores.

Standardization Study

The norms, the raw scores, and corresponding percentiles of the TDL are in Table 3. The large amount of items in the test (36) enabled the clear differentiation and classification of the reading performance of students. The data showed that the instrument was reliable in assessing the reading ability of students ranging from weak to high performance up to the 4th grade. The ceiling effect found in the 5th grade shows that the TDL has limitations in the discrimination the reading performance of students at advanced levels of schooling.

Table 3. *TDL's Percentile Norms for Public State Schools (Raw Scores) According to the Child's School Grade and Chronological Age*

Reading performance	Percentile	School Grade				Chronological age in years				
		2nd	3rd	4th	5th	7	8	9	10	11
Disability	7	0	2	6	10	0	2	3	9	12
	10	0	4	8	12	0	3	6	10	13
	15	3	7	11	14	0	5	10	13	17
Low	25	5	10	13	18	4	6	12	16	21
	30	6	11	15	19	5	7	13	17	23
Average	40	7	13	17	22	7	9	15	19	25
	50	8	15	19	24	8	11	17	22	26
	60	10	17	22	26	10	13	19	24	27
Above average	70	12	19	25	27	12	15	22	26	29
	80	13	22	26	30	13	17	25	29	32
High	90	16	26	29	33	15	23	28	31	35
	95	22	28	32	35	18	26	30	34	36

An analysis of covariance (ANCOVA), with school grade and chronological age as covariates, showed no difference between genders for the TDL ($F(1, 480) = 1.48$, $MSE = 56.2$, $p < .23$). For this reason, the percentile norms of the present study were not split into male and female, as in the Piérart and Grégoire (2004) and in the Sucena and Castro (2010) standardization studies. Consistent with the results of Piérart and Grégoire (2004), there was a

tendency for females to outperform males in the 5th grade ($F(1, 139) = 2.78$, $MSE = 61.8$, $p = 0.97$). Although chronological age could not be used as a covariate, a meta-analysis by Sucena and Castro (2010) demonstrated that the gender difference was concentrated in the 4th grade ($F(1, 171) = 6.6$, $MSE = 29.2$, $p < .011$) and not in all grades as previously reported.

The distinct groups delineated by reading performance in the standardization were supported by a two-step cluster analysis, which suggested that a five-class solution had a good fit-model for all the school grades (see Figure 1). An univariate analysis of variance (ANOVA), with a Bonferroni correction, confirmed that all five groups had significant differences in TDL scores (2nd grade: $F(4, 101) = 499.0$, $MSE = 1.8$, $p < .001$; 3rd grade: $F(4, 102) = 380.3$, $MSE = 4.0$, $p < .001$; 4th grade: $F(4, 125) = 567.7$, $MSE = 3.5$, $p < .001$; 5th grade: $F(4, 136) = 446.4$, $MSE = 4.6$, $p < .001$).

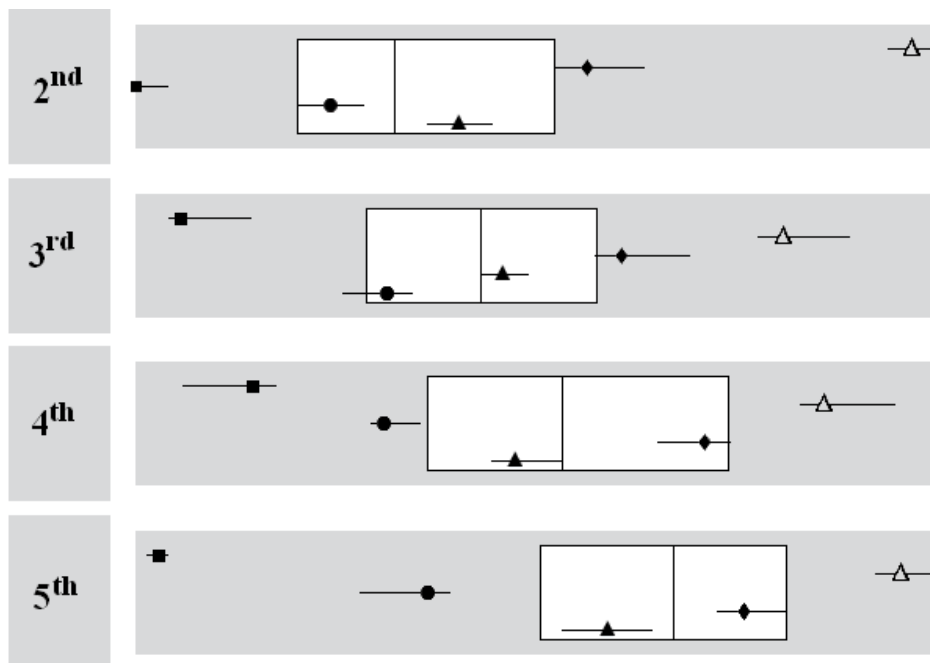


Figure 1. Box plot of the five subgroups identified by cluster analysis based on the score of TDL by school grade

■: Reading disability; ●: Low performance; ▲: Average performance; ◆: Above average performance; △: High performance

The present study found illiterate students (< 3 points on the TDL) in the 2nd (15%), 3rd (8%), 4th (4%) and 5th grade (3%). This data is in agreement with the literacy assessment program [Programa de Avaliação da Alfabetização (PROALFA), 2013], which found that 8.7% of students in the 3rd grade were classified as illiterate. The PROALFA assessment was performed in 94% of public state schools in Belo Horizonte during the period as the data collection of the present study.

Table 4 presents the cross-cultural norms for the L3, TIL, and TDL. It includes the norms for a sample of 253 students from the 5th grade (Cunha, 2010). As these norms are not statistically different to those found by Sucena and Castro (2010), the results of both studies will be merged in the subsequent cross-cultural comparison of TIL. The Belgium sample had lower means than the Portuguese sample in all comparable school grades (3rd, 4th, and 5th grade). Compared with the Brazilian sample, the Belgium scores were lower in the 3rd grade, but equivalent in the 4th and 5th grades. The norms from Portugal were statistically higher ($p < .01$) than the Brazilian in all the school years compared (2nd–5th grade).

Table 4. *Cross-cultural Comparison With the Normative Data From the L3 test, TIL and TDL*

Test	Reference	School Grade				
		2nd	3rd	4th	5th	6th
L3	Piérart and Grégoire (2004)*		N = 765	N = 723	N = 791	N = 710
			M = 13,1	M = 18,8	M = 23,1	M = 32,5
			SD = 6.7	SD = 6.8	SD = 6.4	SD = 5.7
TIL	Sucena and Castro (2010)* merged with Cunha (2010)	N = 170	N = 186	N = 173	N = 338	
		Age = 7.5	Age = 8.6	Age = 9.6	Age = 11.1	
		M = 13.8	M = 19.3	M = 25.5	M = 28.7	
		SD = 5.1	SD = 4.9	SD = 5.4	SD = 5.9	
TDL	Present work	N = 106	N = 107	N = 130	N = 141	
		Age = 8.1	Age = 8.9	Age = 9.8	Age = 10.8	
		M = 9.1	M = 15.3	M = 19.3	M = 23.2	
		SD = 5.9	SD = 7.8	SD = 8.0	SD = 7.9	

Note. *Male and female values were combined to create a general score.

Discussion

The objective of this study was to validate and standardize the Reading Performance Screening Test (TDL). As expected, the TDL had good correlations with all the comparison instruments. Among these results, the TDL's strongest correlation was found with the accuracy rate of word recognition ($r = .840$). This is in agreement with D'Hondt and Leybaert (2003), who found a significant correlation between the L3 test with a timed lexical decision task ($r = .65$, $p < .001$). These results are logical because all these tasks are timed measures, which require quick identification of words, a process that is indispensable to freeing cognitive resources for the understanding of sentences (Dehaene, 2012; Ehri, 2010; Share, 1995).

As the TDL does not evaluate specific linguistic features, but incorporates the lexical word recognition and semantic component of reading, the high correlation with the general reading composite can be considered the most important result of the current study. This composite integrates three reading measures into one, through a dimension reduction technique, and enables representation of reading performance in a single robust variable.

As expected, general cognitive ability played a moderate role in TDL scores ($r = .50$). This value is within the range (.36 to .68; $M = .49$) found by Carver (1990) on the CPM and the National Reading Standards test with students in Grades 2–12. The abilities measured by the CPM seem to have a moderate and consistent relationship to the reading ability.

Concerning the psychiatric behaviors, all signs of behavioral problems had a negative effect, although weak, on TDL scores, which is in agreement with the literature; a negative association between indicators of externalizing behavioral and school outcomes was expected (Kristoffersen, Obel, & Smith, 2014). In contrast, prosocial behavior was beneficial to the reading performance of students.

Cross-cultural Comparison

This subsection compares the norms for the L3 (Belgium: Piérart & Grégoire, 2004) with those of both the TIL [Portugal: a merge of the results of Cunha (2010) and Sucena & Castro (2010)] and the TDL (Brazil: results of the present study). Further, the norms of the two Portuguese versions (the TIL and TDL) were also compared. It is important to note that the compulsory education starts at age six in all three countries.

It was not anticipated that the L3 would have statistically lower means than the TIL in all comparable school grades. Interestingly, our expectation was that the Portuguese children would have a slightly worse or equivalent mean as the Belgium students. These predictions were driven by two reasons. First, the syllabic complexity and orthographic depth of European Portuguese is close to French (Seymour, Aro, & Erskine, 2003), and, in principle, it is logical

to anticipate an equivalent degree of difficulty to learn to read in both orthographies. Portuguese and French children read a similar number of words (75% and 79%, respectively) and pseudowords (75% and 85%) (Seymour, Aro, & Erskine, 2003). Second, the Belgium education system ranks among the top worldwide, as its students' mean performance on the PISA reading scale was statistically significant above the Organization for Economic Cooperation and Development (OECD) average at the time of the L3 standardization, whereas the scores of Portuguese students ranged from below to within OECD averages (OECD, 2003). Thus, before the results of this cross-cultural comparison, we argued that the TIL is perhaps a less demanding test than the L3. This assertion can be substantiated on the ground of two observations about the TIL: less control of the linguistic variables in the test and its size.

Sucena and Castro (2010) did not present: a) a clear classification of the distractors of the L3 test; b) the logic for the adaptation of the original sentences, target words, and distractors; c) a reverse translation (from European Portuguese to French), or d) a random sample. Therefore, we used the L3, not the TIL as a reference for the Brazilian adaptation. The TIL is shorter in terms of the number of letters than L3 (3118 vs. 3598 letters), which is equivalent to a test 13% smaller. In this time-controlled measure, the letter difference may matter and should have been taken into account in the instruments designed to allow cross-cultural comparisons.

The result of the Belgium and Brazilian comparison was expected, as they differed in the 3rd grade, but were equivalent in the 4th and 5th grades. As Brazilian Portuguese orthography is much more regular (Pinheiro, 2011; Pinheiro & Roth-Neves, 2001; Scliar-Cabral, 2003) than French, a better performance on TDL than L3, at least in the initial steps of literacy (namely, in the 1st through 3rd grades) could be expected. This advantage may have been lost due to the low quality of the Brazilian Educational System, demonstrated in all international evaluations. For example, the scores of Brazilian students were well below the OECD's average (OECD,

2012). Thus, it is not surprising that the Belgium sample caught up with the Brazilian sample. The similarity of results means that the two tests may be equivalent in terms of difficulty.

The higher results in Portugal compared to Brazil, in all the school years compared, is contrary to the orthographic depth theory that asserts a more transparent orthography leads to faster literacy learning (Seymour, Aro, & Erskine, 2003). Brazilian Portuguese is more regular than European Portuguese; therefore, it should be easier to learn. In addition, the Brazilian children were an average of 4 months older than the children in the TIL standardization; the obtained results are contrary to the expectations of human development and maturation.

However, as the TIL is a more lenient test than the TDL, the gap between the two versions may have been amplified, leading to “false” better performance by the Portuguese children. On the other hand, the observed differences may be due to **the Brazilian educational method or the socio-economic disadvantage**. Our sample was entirely composed of children from public schools, which have lower performance and more children designated low socioeconomic status (SES) than private schools. Research has shown that low SES limits the access of children to various types of cultural resources (Duncan & Seymour, 2000; Soares, 2004). Duncan and Seymour found that children from different socioeconomic levels had significant differences in letter name recognition tasks and reading aloud of real words and pseudowords. Such differences are indicative of a delay of at least one year in the acquisition of reading skills for children of low socioeconomic status.

It is important to point out that these socioeconomic effects are also found in Portugal, where a large study conducted by Feitosa, Matos, Del-Prette, and Del-Prette (2005) found a significant correlation ($r = .033$) between academic performance and socioeconomic level. However, Silva (2011) found only a correlation tendency between TIL and the children’s socioeconomic background (possibly due to a small sample size).

The Brazilian government's "National Pact for Literacy in the Correct Age" (Ministério da Educação, 2012) represents a commitment to ensure that all children become literate by the age of eight years, which is equivalent to the end of the 3rd year of elementary school. Although this program reflects an interest in changing the *status quo*, it is very permissive, as the present data demonstrated that the state educational system has not been completely successful with its literacy program, based on the number of illiterate children in the 4th and 5th grades. This result also demonstrates the need for the identification of reading problems by the 2nd grade to prepare for educational interventions.

Finally, according to a recent evaluation by the Public Basic Education Assessment Program [Programa de Avaliação da Rede Pública de Educação Básica (PROEB), 2013] 21.3% of the 5th grade students from state schools in Minas Gerais have low proficiency in Portuguese and 32.2% have an intermediate proficiency. This means that standardization studies are necessary for private school systems to provide a more rigorous norm for TDL.

Conclusion

The TDL is a valid instrument to assess the global reading competence of students' school year (2nd through 5th grades) and chronological age (7 to 11 years). Due to the psycholinguistics controls introduced, the TDL is equivalent to the L3 and more robust than the TIL. Thus, it can be used for collective screening or individual clinical administrations. These features make the TDL an important psychometrically standardized measure to assess the Criterion B for Specific Learning Disorder (particularly with the specifier for impairment in reading, which is also referred as Dyslexia) in the DSM-5, which mandates an academic skill substantially and quantifiably below those expected for the individual's chronological age.

Conflict of Interest

The authors declare no conflicts of interest.

References

- Alegria, J., Domínguez, A. B., & Straten, P. (2009). ¿Cómo leen los sordos adultos? La estrategia de palabras clave. *Revista de Logopedia, Foniatría y Audiología*, 29(3), 195–206.
- Angelini, A. L., Alves, I. C. B., Custódio, E. M., Duarte, W. F., & Duarte, J. L. M. (1999). *Matrizes Progressivas Coloridas de Raven: Escala Especial*. São Paulo: CETEPP.
- Carreira-Perpiñán, M. Á. (1997). *A Review of Dimension Reduction Techniques*. Technical Report. Department of Computer Science, University of Sheffield.
- Carrillo, M. S., & Marín, J. (2009). Test de Eficiencia Lectora – TECLE. In A. Cuadro, D. Costa, D. Trias, & P. Ponce de León (Eds.), *Evaluación del nivel lector: Manual técnico del test de eficacia lectora (TECLE)*. Uruguay: Prensa Médica Latinoamericana.
- Carver, R. P. (1990). Intelligence and reading ability in grades 2-12. *Intelligence*, 14(4), 449–455.
- Cogo-Moreira, H., Ploubidis, G., De Avila, C., Mari, J., & Pinheiro, A. M. V. (2012). EACOL (scale of evaluation of reading competency by the teacher): Evidence of concurrent and discriminant validity. *Neuropsychiatric Diseases and Treatment*, 8, 443–454.
- Colin, S., Leybaert, J., Ecalte, J., & Magnan, A. (2013). The development of word recognition, sentence comprehension, word spelling, and vocabulary in children with deafness: A longitudinal study. *Research in Developmental Disabilities*, 34, 1781–1793.
- Cramer, D., & Howitt, D. (2004). *The Sage Dictionary of Statistics: A practical resource for students in the social sciences*. Thousand Oaks: Sage.
- Capellini, S. A., Oliveira, A. M., & Cuetos, F. (2012). *PROLEC: Provas de Avaliação dos Processos de Leitura* (2nd ed.). São Paulo: Casa do Psicólogo.
- Cunha, J. P. C. (2010). *Análise e Intervenção em Problemáticas Específicas de Leitura* (Master's thesis). Retrieved from <http://run.unl.pt/handle/10362/5138>

- Cury, C. R., & Golfeto, J. H. (2003). Strengths and difficulties questionnaire (SDQ): a study of school children in Ribeirão Preto. *Revista Brasileira de Psiquiatria*, 25(3), 139–145.
- D'Hondt, M., & Leybaert, J. (2003). Lateralization effects during semantic and rhyme judgement tasks in deaf and hearing subjects. *Brain and Language*, 87, 227–240.
- Dehane, S. (2012). *Os neurônios da leitura: Como a ciência explica a nossa capacidade de ler*. Porto Alegre: Penso.
- Duncan, L. G., & Seymour, P. H. K. (2000). Socio-economic differences in foundation level literacy. *British Journal of Psychology*, 91, 145–166.
- Ehri, L. C. (2010). Development of Sight Word Reading: Phases and Findings. In M. J. Snowling & C. Hulme (Eds.). *The science of reading: A handbook* (pp. 362–378). Oxford: Blackwell Publishing.
- Feitosa, F. B., Matos, M. G., Del-Prette, Z. A. P., & Del-Prette, A. (2005). Suporte social, nível socioeconômico e o ajustamento social e escolar de adolescentes portugueses [Social support, socioeconomic level, and school-social adjustment of Portuguese adolescents]. *Temas em Psicologia*, 13(2), 129–138.
- Génard, N., Mousty, P., Content, A., Alegria, J., Leybaert, J., & Morais, J. (1998). Methods to establish subtypes of developmental dyslexia. In P. Reitsma & L. Verhoeven (Eds.), *Problems and interventions in literacy development* (pp. 163–176). Netherlands: Kluwer Academic Press.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586.
- Goodman, R. (2005). Questionário de Capacidades e Dificuldades (SDQ-Port). Retrieved from www.sdqinfo.com

- Hoaglin, D. C., & Iglewicz, B. (1987). Fine tuning some resistant rules for outlier labeling. *Journal of American Statistical Association*, 82, 1147–1149.
- Kristoffersen, J. H. G., Obel, C., & Smith, N. (2014). Gender differences in behavioral problems and school outcomes. *Journal of Economic Behavior & Organization*. Advance online publication. doi:10.1016/j.jebo.2014.10.006
- Leybaert, J. (2000). Phonology acquired through the eyes and spelling in deaf children. *Journal of Experimental Child Psychology*, 75, 291–318.
- Lobrot, M. (1967). *Batterie pour mesurer la lecture et l'orthographe, ORLEC*. Beaumont/Oise: Bureau d'études et de recherches.
- Lobrot, M. (1980). *Lire avec e ´preuves pour e ´valuer la capacite ´ de lecture (D-OR-LEC)*. Paris: Editions ESF.
- Marín, J. & Carrillo, M. S. (1999). *Test Colectivo de Eficacia Lectora (TECLE)*. Published Manuscript. Departamento de Psicología Básica y Metodología. Universidad de Murcia. Available from www.pediatrasandalucia.org/Pdfs/TestEficienciaLectora.pdf
- Ministério da Educação, Brasil. (2012). *Manual do pacto: Pacto pela Alfabetização na Idade Certa: o Brasil do futuro com o começo que ele merece*. Brasília, DF.
- Mousty, P., & Leybaert, J. (1999). Evaluation des habilités de lecture et d'orthographe au moyen de BELEC: données longitudinales auprès d'enfants francophones testés en 2^{ème} et 4^{ème} années. *Revue Européenne de Psychologie Appliquée*, 49, 325–342.
- Mussolin, C., Mejias, S., & Noël, M. P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition*, 115, 10–25.
- Organization for Economic Co-operation and Development (OECD). (2003). *Programme for International Student Assessment. Results: Executive Summary*. Available from

<http://www.oecd.org/edu/school/programme-for-international-student-assessment-pisa/34002454.pdf>

- Organization for Economic Co-operation and Development (OECD). (2012). *PISA 2012 Results in Focus: What 15-year-olds know and what they can do with what they know*. Available from <http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf>
- Pasquali, L., Wechsler, S. M., & Bensusan, E. (2002). Matrizes Progressivas do Raven Infantil: um estudo de validação para o Brasil. *Avaliação Psicológica*, 1(2), 95–110.
- Piérart, B., & Grégoire, J. (2004). Déchiffrer et coprendre: le test de closure en lecture revisité étalonnage belge du L3 de Lobrot. *Le Langage et l'Homme*, 39(2), 87–100.
- Pinheiro, A. M. V. (1996). *Contagem de Frequência de Ocorrência de Palavras Expostas a crianças na faixa pré-escolar e séries iniciais do 1º grau*. São Paulo: Associação Brasileira de Dislexia.
- Pinheiro, A. M. V. (2011). Transparência ortográfica e o efeito de retroalimentação fonológico grafêmica: implicações para a construção de provas de reconhecimento de palavras. In L. M. Alves, S. Capellini, R. Mousinho (Eds.), *Dislexia: Novos temas, novas perspectivas* (pp. 131–146). Rio de Janeiro: Editora Wak.
- Pinheiro, A. M. V. (2013). Prova de Leitura e de Escrita de palavras. Relatório Técnico Final aprovado pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG (FAPEMIG). Número do processo: APQ-01914-09.
- Pinheiro, A. M. V., & Rothe-Neves, R. (2001). Avaliação cognitiva da leitura: as tarefas de leitura em voz alta e ditado. *Psicologia: Reflexão e Crítica*, 14(2), 399–408.
- Primi, R., Couto, G., Almeida, L. S., Guisande, M. A., & Miguel, F. K. (2012). Intelligence, age and schooling: data from the Battery of Reasoning Tests (BRT-5). *Psicologia: Reflexão e Crítica*, 25(1), 79–88.

- Programa de Avaliação da Alfabetização (PROALFA). (2013). *SIMAVE – Sistema Mineiro de Avaliação da Educação Pública*. Secretaria de Estado de Educação de Minas Gerais.
- Programa de Avaliação da Rede Pública de Educação Básica (PROEB). (2013). *SIMAVE – Sistema Mineiro de Avaliação da Educação Pública*. Secretaria de Estado de Educação de Minas Gerais.
- Reybroeck, M., & Hupet, M. (2009). Acquisition of number agreement: effects of processing demands. *Journal of Writing Research, 1*(2), 153–172.
- Rousselle, L., & Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition, 102*(3), 361–395.
- Saur, A. M., & Loureiro, S. R. (2012). Psychometric properties of the Strengths and Difficulties Questionnaire: a literature review. *Estudos de Psicologia (Campinas), 29*(4), 619–629.
- Scliar-Cabral, L. (2003). *Princípios do sistema alfabético do português do Brasil*. São Paulo, Editora Contexto.
- Serniclaes, W., Heghe, S., Mousty, P., Carré, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology, 87*, 336–361.
- Seymour, P. H. K., Aro, M., & Erskine, J. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143–174.
<http://dx.doi.org/10.1348/000712603321661859>
- Share, D. L. (1995) Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition, 55*, 151–218.

- Silva, J. C. P. (2011). *A Influência do nível socioeconómico na leitura*. Graduation Project. Porto: Universidade Fernando Pessoa, Escola Superior de Saúde. Available from <http://hdl.handle.net/10284/2843>
- Soares, J. F. (2004). O efeito da escola no desempenho cognitivo de seus alunos. *REICE – Revista Eletrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 2(2), 83–104.
- Spiegel, M. R., Schiller, J. J., & Srinivasan, A. L. (2000). *Schaum's outline of probability and statistics*. New York: McGraw-Hill.
- Sucena, A., & Castro, S. L. (2010). *Aprender a Ler e Avaliar a Leitura. O TIL: Teste de Idade de Leitura* (2nd ed.). Coimbra: Almedina.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Vilhena, D. A., Sucena, A., Castro, S. L., & Pinheiro, A. M. V. (2014). Reading Performance Screening Test: An adapted version of Lobrot's L3 test for Brazilian Portuguese. Manuscript submitted for publication.

ARTIGO 3**EACOL (Scale of Evaluation of Reading Competence by the Teacher): Modification and
Validation**

Douglas de Araújo Vilhena* and Ângela Maria Vieira Pinheiro*

*Universidade Federal de Minas Gerais, Brazil

Abstract

This study validates, standardizes, and improves the EACOL, a tool for teachers to assess the Portuguese-language reading (silent and aloud) of 2nd-to-5th-grade students. Modifications made were: a) replacement of “Yes”/”No” answers by “True”/”False”; b) addition of answers “Sometimes” and “I do not know”; c) removal, addition, and revision of items; d) selection of the best scoring rubric. The instrument presented high internal consistency and moderate-to-strong correlations with all seven reading variables; cluster analysis suggested the existence of three proficiency groups (poor/average/good readers). Discussion of discriminant validity is provided. We hope to offer to Portuguese-speaking researchers a validated instrument to indirectly assess the reading ability of schoolchildren, and to set out a model that can be adapted to other contexts.

Keywords: reading assessment, child assessment, Portuguese language, teacher scale, alternative assessment, scoring rubric

1. Introduction

According to the Literacy Initiative for Empowerment (LIFE; United Nations Educational, Scientific and Cultural Organization [UNESCO], 2007), education is a human right and a public good that enables access to information about health, the environment, the world of work and, most importantly, how to learn throughout life. This idea must take on particular importance in Brazil, in which only 56.1% of children are fully literate at 8 years of age (Todos pela Educação, 2013) and 11% of young people aged 15–24 remain functionally illiterate (Instituto Paulo Montenegro, 2011).

Given this situation, a proactive approach is needed. Nothing justifies simply waiting for students to fail, as the focus of literacy education should be on the prevention of reading problems rather than on remedial intervention. Early screening for reading difficulties can be appropriately done by elementary school teachers, who are undeniably one of the most important sources of information about their students. According to Snowling, Duff, Petrou, Schiffeldrin, and Bailey (2011), who examined the predictive validity of some scales measuring reading behavior, when criterion-referenced assessments are made available to teachers, their evaluations of the reading skills of their students can be as good as those of most formal tests.

In recent years, there has been a worldwide emphasis on constructing and monitoring teacher and pupil educational standards (Dobson, 2012; Gove & Cvelich, 2011). In Brazil and in many other countries, there is a lack of instruments with the validity and precision to guide teachers in an initial categorization of the reading abilities of their students. The development of the Scale of Evaluation of Reading Competence by the Teacher (in Portuguese, Escala de Avaliação da Competência em Leitura pelo Professor, or EACOL) is an initiative to fill this gap (Pinheiro & Costa, 2012, 2015), but previous studies identified issues indicating that the

scale needed revision. The goal of this paper is to make several structural improvements in the EACOL in response to these issues and to conduct subsequent validation.

1.1 The EACOL

Pinheiro and Costa (2005) began the development of the EACOL with a validation by the judgment of specialists of a set of descriptors of good, average and poor *Reading Aloud* (RA) and *Silent Reading* (SR) behaviors that could be recognized by the teacher. RA items measure speed and accuracy in word recognition, prosody, and comprehension; whereas SR items measure comprehension and the capacity for synthesis. After this procedure, two scales were created: a) *Form A*, with 23 items for 2nd-graders (in elementary school), who are at or near the beginning of the literacy process, with an average age of 7 years; and b) *Form B*: with 27 items for students from 3rd to 5th grade, at the later stage of literacy learning and also for readers already literate, with an approximate age of 8–11 years.

Although the EACOL has already been submitted to two previous validation studies, some important methodological and structural issues remain. In the first study, conducted by Cogo-Moreira, Ploubidis, De Avila, Mara, and Pinheiro (2012), using the statistical Latent Class Analysis method, the three types of readers expected by the authors of the EACOL (good, average, and poor readers) were found. However, there was an overlap between two items of the scale—*Reads too slowly or too quickly* and *Reads words correctly*—suggesting that they should be revised or removed. Additionally, this study assessed discriminant validity based on psychiatric behaviors displayed and so-called non-verbal intelligence, an approach that goes against a number of theoretical and empirical studies, since both variables are correlated with the reading ability (e.g., Baker & Ireland, 2007; Carver, 1990; Maughan & Carroll, 2006).

In the other study, Lúcio and Pinheiro (2013) compared teachers' judgments with the performance of their students on a Reading Aloud Word Task (Pinheiro, 2007). In general, they found moderate correlations, with weak results mainly in the Silent Reading subscale. This

suggested a necessity of establishing concurrent validation for the EACOL SR subscale by means of a measure that assesses this type of reading rather than Reading Aloud. Another issue detected was a decline in the correlation in the good reading ability group.

In both previous validation studies, there were a significant number of items left empty by the teachers. This could have resulted from the dichotomous nominal level of measurement of the instrument, with only “Yes” and “No” answers, as in this case a teacher may be prone to waive an answer if he or she is not pleased with either alternative (Pinheiro, 2013a). In addition, when faced with a binary choice, the respondents favors positive alternatives rather than negative ones (Emmerich, Enright, Rock, & Tucker, 1991). Thus, in an attempt to obtain more control over the answers given by teachers and to avoid the problems associated with binary options, Pinheiro (2013a) suggests the addition of the alternative “I do not know.”

Another observation refers to the intelligibility of the items for the average reader category. Many of these items contain the word “sometimes,” which can confuse and reduce the accuracy of the teacher’s assessment (Lúcio & Pinheiro, 2013). Examples of such items are: *Sometimes makes mistakes when reading “new” words* and *Sometimes reads and cannot retell what was read*. We reasoned that the inclusion of a “Sometimes” option within the alternatives given to teachers could be beneficial, even though this alteration required a further change in the structure of the scale: the exclusion of all item descriptors pertaining to the average reader.

Finally, again inspired by studies evaluating the reliability of multiple-choice answers (e.g., Verbič, 2012), we replaced the options “Yes” and “No” with “True” and “False” to avoid misinterpretation of items making negative statements. For example, on the item *Not always able to identify the subject from the title and vice versa*, while a “Yes” answer indicates a poor reader, a “No” answer indicates a good reader. In such cases, the teacher may erroneously

assign a “Yes” to a good performance or a “No” to a poor performance, which would lead to an inaccurate judgment of the child’s ability.

Therefore, to summarize, the EACOL underwent the following modifications: a) replacement of “Yes” by “True” and “No” by “False”; b) replacement of the binary option for answers by four alternatives: “True,” “False,” “Sometimes,” and “I do not know”; c) exclusion of some items due to the new response format; d) addition and revision of other items; and e) identification and selection of the best scoring rubric. After all these changes, new validation, precision, and standardization studies are required, as the latent structure of the scale could have been modified.

2. Method

2.1 Participants

To evaluate whether the teacher’s judgment is as reliable as direct reading assessment, the cognitive functions of 2nd-to-5th-graders were evaluated to provide concurrent validity (see Table 1 for the pupils’ sociodemographic distribution). The sample (452 students and 72 teachers across 8 state schools) was gathered from November to December 2013. The institutions, arbitrarily chosen from a document provided by the Minas Gerais State Secretary of Education, were distributed over five of the nine school districts in Belo Horizonte.

Each teacher was asked to answer the EACOL and a screening behavioral scale for six students only. These pupils, randomly selected, performed a test battery composed by the Reading Performance Screening Test, a Word and Pseudoword Reading Task, Reading Comprehension Test, and a general cognitive ability test.

Table 1. *Breakdown of the student sample according to age, sex, and grade*

Age in years	Sex		Grade				Total
	Male	Female	2 nd	3 rd	4 th	5 th	
7	24	15	39				39
8	48	61	57	51	1		109
9	61	48		41	65	3	109
10	62	76		3	57	78	138
11	26	28		1		53	54
12	1	0				1	1
13	2	0				2	2
Total	224	229	96	96	123	137	452

Schools, teachers, students and their guardians signed an informed consent form for the research. The assessments were administered during school hours, in a quiet room in the institution. All participants provided informed consent, and the Ethical Committee from the Federal University of Minas Gerais approved the study (Certificado de Apresentação para Apreciação Ética [Certificate of Appreciation Presentation to Ethics; CAAE]: 17754514.6.0000.5149).

2.2 Instruments administered to the teachers

The EACOL used in the present study, as in all the previous studies, is composed of two forms (A and B) that differ in their number of items and in part of their content. Form A consists of 15 items and Form B of 21 items (against 23 and 27 items, respectively, in the previous version of the instrument). This decrease in the number of items is due to the exclusion of the items for the category “average reader.” After all items are the alternative answers “True,” “False,” “Sometimes,” and “I do not know.” To guarantee the internal consistency of

the instrument, two criteria were established to control for incongruence and/or unjudgeability on a given scale: a) opposing items answered more than twice, and b) the presence of four or more “I do not know” responses. Either of these led to the exclusion of that scale from the sample.

Child behavior was assessed by the Strengths and Difficulties Questionnaires (SDQ), which is a brief behavioral screening questionnaire for 4–16-year-olds (Goodman, 1997; Cury & Golfeto, 2003; Saur & Loureiro, 2012). This study used the single-sided Brazilian version, without impact supplement, with scoring for teachers (Goodman, 2005). This instrument has 25 items divided into 5 scales: *emotional symptoms* (anxiety/mood), *conduct problems* (aggression/delinquency), *hyperactivity/inattention*, *peer relationship problems* (withdrawn/social problems), and *prosocial behavior* (empathy/positive relations).

2.3 Instruments administered to the students

The Word Reading Task (WRT) and the Pseudoword Reading Task (PWRT) are Reading Aloud instruments each consisting of 88 words and 88 pseudowords (Pinheiro, 2013b; Cogo-Moreira et al., 2012). The psycholinguistic variables for the words were a) *frequency of occurrence* (high vs. low), b) *bidirectional regularity* (regular and irregular words according to grapheme-to-phoneme correspondence and vice versa), and c) *length* (short, medium, and long words). The pseudowords were constructed with the same orthographic structures and stimulus length used in the word task. On both instruments, two measures were used: a) *accuracy* (total number of correctly read words or pseudowords) and b) *accuracy rate* (total number of correct words or pseudowords read per minute).

The Reading Performance Screening Test (TDL) was used to evaluate the Silent Reading efficiency of children from 7 to 11 years old. The test takes into account lexical word recognition, syntactic competence, comprehension and speed (Vilhena & Pinheiro, 2014; Vilhena, Sucena, Castro, & Pinheiro, 2014). It consists of 40 incomplete and isolated sentences,

each followed by five words as alternative fill-in-the-blank answers. The child's task is to select as quickly as possible (taking up to 5 minutes) the best word to give meaning to each sentence.

Another instrument used to evaluate the Silent Reading was the Text Reading Comprehension subtest, which is part of the PROLEC (Provas de Avaliação dos Processos de Leitura [Reading Processes Assessment Battery]; Capellini, Oliveira & Cuetos, 2012). It consists of four short texts to investigate students' ability to extract meaning and integrate it with prior knowledge. Each text has 4 questions (literal and inferential), resulting in 16 questions overall.

General cognitive ability was measured using Raven's Coloured Progressive Matrices Test (CPM) (Angelini, Alves, Custódio, Duarte, & Duarte, 1999). It evaluates analogic reasoning, or the ability to infer relations between objects or elements (Pasquali, Wechsler, & Bensusan, 2002). It is used mainly for children between 5 and 11 years, and consists of 36 items divided into three sets of 12 (A, Ab, B) arranged in inter- and intraset according to increasing difficulty. The task is presented as a puzzle game, with the aim being to select the best option among six alternatives printed beneath.

2.4 Statistical analyses

All analyses were performed using IBM SPSS Statistics version 21.0. Due to the diversity in the item structures, all data were transformed to represent only a Likert-type scale from negative to positive.

A hypothetical-deductive method using a Pearson bivariate correlation with all the instruments was applied to determine which was the best scoring rubric for the alternatives of each item from the EACOL. Four hypotheses were tested: a) bad reading: 0, average: 1, good: 2; b) bad reading: 0, average: 2, good: 3; c) bad reading: 0, average: 1, good: 3; d) bad reading: 0, average: 0, good: 2. The answer "I do not know" was assigned the same score as an average reader's answer.

Cronbach's alphas were calculated to estimate the internal consistency reliability associated with the scores of each item on the EACOL's Forms A and B. A hypothetical-deductive method can confirm if the removal of any item can alter the alpha and the concurrent validity correlations.

As EACOL evaluates reading competence as a whole, dimension reduction by principal component analysis (Carreira-Perpiñán, 1997) was used to incorporate all four reading instruments into a robust reading measure, from here on called the General Reading Composite (General RC). A reliability analysis indicated the use of the raw scores from the PROLEC, TDL, Word Reading Task accuracy rate, and Pseudoword Reading Task accuracy. This integration of measures enables us to represent the child's reading performance with a single variable.

A two-step cluster analysis was used to verify the number of mutually exclusive latent groups in the sample. The only variables used were the score for each item in EACOL. This method is a scalable cluster analysis algorithm designed to handle large data sets in two steps: 1) pre-cluster the cases into many small sub-clusters; 2) cluster these sub-clusters into the desired number of clusters. The log likelihood distance measure was used, with subjects assigned to the cluster leading to the largest likelihood. The Bayesian information criterion (BIC) was established to compare the number of latent classes, a comparison in which small values correspond to better fit. Differences in the sample were compared according to cluster membership using a univariate Analysis of Variance (ANOVA) test. For all tests performed, the significance level was set at .05, two-tailed.

3. Results

3.1 Item revision

Due to the addition of the alternative "Sometimes," the following eight items, descriptors of the average reader, were removed in both Form A and Form B: a) *Sometimes*

reads and cannot retell what was read; b) Reads too slowly or too quickly; c) Sometimes makes mistakes when reading “new” words; d) Sets the tone of interrogation and/or exclamation only in the word that precedes the punctuation mark; e) Slows the rhythm of reading when “new” words are encountered, needing to spell them out; f) Not always able to identify the subject from the title and vice versa; g) Does identify characters and places, but has some difficulty identifying main ideas without a second reading; and h) Has some difficulty in orally summarizing what was read.

The two items that showed poor discrimination in Cogo-Moreira et al. (2012) were also removed, because one is a descriptor of an average reader (*Reads too slowly or too quickly*) and the other was rather vague (*Reads words correctly*). Finally, the last excluded item was a descriptor of a poor reader (*Says “I do not know” when encounters a new word*), since there is another item in the scale that deals with reading of new words and to avoid confusion with the new alternative answer “I do not know.”

In contrast to these 10 removed items, 5 others were added (one in Form A and the remainder in Form B). This was thought to be necessary to increase the number of descriptors of the ability of the readers and to maintain the power of the scale. The descriptor of poor reading *Reads with difficulty “known” words* was added to Form A (in the Reading Aloud subscale), the following items were added to Form B (also in the Reading Aloud subscale): a) *Reads clearly, without “stumbling” or “swallowing” syllables. Someone who hears can understand what is being read;* and b) *Has great difficulty in Reading Aloud.* As for the Silent Reading subscale of Form B, the additions consisted of a) *Reads without pronouncing words or without moving the lips, only moving the eyes;* and b) *Cannot read without movements of the lips or without pronouncing the words.*

Finally, the item *Reads “new” and invented words quickly* was changed into *Reads “new” words correctly*. The omission of “invented words” was motivated by the fact that

pseudowords are rarely presented to students in school. Equally, the alteration of *quickly* into *correctly*, was motivated by the expectation that more important and urgent than reading the “new words” quickly was reading them correctly.

3.2 Validation

On the selection of the scores for EACOL, the strongest correlations were with the first hypothesis (the first rubric). This was the hypothesis under which predictors of poor readers are scored zero, predictors of good readers score two points, and the alternatives “Sometimes” (=average readers) and “I do not know” are scored one point (see Appendix).

In Form A, the Cronbach’s Alpha for the Reading Aloud subscale was .891, with the corrected item–total correlations indicating that the alpha would lose its power if any item were removed. On the other hand, for the Silent Reading subscale, the removal of item 5 (*Does not identify characters, places, or main ideas*) increases the alpha from .915 to .919. This exclusion recommendation was confirmed by the consistent weak correlations of item 5 ($r \approx .244$) with all reading measures. Finally, the total score (sum of both subscales minus the aforementioned item 5) has an alpha of .935, demonstrating the strong internal consistency reliability of EACOL’s Form A. In the further analysis of Form A, item 5 will not be considered.

The same internal validity test was performed on Form B. The subscales Reading Aloud and Silent Reading demonstrated strong Cronbach’s alphas (.940 and .933, respectively), with a loss in alpha with the removal of any item. The alpha of the Total score was .958, demonstrating that Form B also has a high internal consistency.

For concurrent validity, to attest to what extent the evaluations of teachers agree with the actual performance of children, correlations were calculated between the scores of EACOL and all reading measures (see Table 2).

Table 2. *Pearson correlation between EACOL, reading, general cognitive ability, and behavior*

Measure	Form A			Form B		
	RA	SL	Total	RA	SL	Total
WRT (accuracy measure)	.637**	.628**	.670**	.513**	.369**	.484**
(accuracy rate)	.650**	.590**	.662**	.593**	.486**	.587**
PWRT (accuracy measure)	.600**	.586**	.631**	.549**	.404**	.522**
(accuracy rate)	.637**	.571**	.647**	.554**	.454**	.548**
<u>TDL</u>	.636**	.583**	.650**	.655**	.539**	.649**
<u>Text comprehension</u>	.511**	.517**	.544**	.516**	.410**	.505**
General reading composite	.710**	.671**	.737**	.703**	.559**	.688**
CPM	.237*	.265**	.263**	.354**	.333**	.370**
Prosocial behavior	.095	.228*	.158	.249**	.274**	.278**
Emotional symptoms	-.307**	-.314**	-.328**	-.240**	-.207**	-.244**
Conduct problems	-.223*	-.314**	-.275**	-.393**	-.353**	-.401**
Hyperactivity/inattention	-.477**	-.477**	-.505**	-.492**	-.413**	-.490**
Peer relationship problems	-.215*	-.285**	-.258*	-.325**	-.290**	-.331**
Total negative behaviors	-.430**	-.479**	-.477**	-.479**	-.415**	-.483**

Note. * $p < .05$ (2-tailed), ** $p < .01$ (2-tailed).

RA: Reading Aloud subscale; SL: Silent Reading subscale; Total: sum of RA and SL; TDL: Reading Performance Screening Test; WRT: Word Reading Task; PWRT: Pseudoword Reading Task; Text comprehension: PROLEC Text Comprehension subtest; CPM: Coloured Progressive Matrices scores. The four underlined variables combined form the General Reading Composite.

Forms A and B had correlation ranges with the reading measures of .511–.738 and .369–.655, respectively. As in previous studies, the Reading Aloud subscale had stronger correlations when compared to Silent Reading.

EACOL incorporates accuracy in word recognition, reading speed, prosody, comprehension and the capacity for synthesis. Therefore, the good correlations found with the General Reading Composite ($r = .559$ to $.737$) can be considered the most important result of the current study, attesting that the teachers, when provided with sound criteria, can come to reliable evaluations of their students' reading ability.

Unlike in Cogo-Moreira et al. (2012), Form B was significantly correlated ($p < .0001$) with CPM ($r = .37$) and with the total score of the SDQ ($r = -.48$). Although it was not tested by Cogo-Moreira et al. (2012), Form A in this study also demonstrated correlations ($p < .0001$) with CPM ($r = .26$) and with all SDQ scales.

As expected, the two-step cluster analysis suggested a good fit-model with the following three classes for Form B: poor ($N = 47$), average ($N = 119$), and good readers ($N = 184$). As seen in Figure 1, a clear three-class group structure is therefore supported, considering both empirical and theoretical elements, with an estimated probability axis scale from 0 (reading disability) to 2 (good reading ability). An univariate Analysis of Variance confirmed that all three groups presented significant distinctions from one another on EACOL Total Scores, $F(2, 347) = 1312.7$, $MSE = 14.4$, $p < .00001$. Again, unlike Cogo-Moreira et al. (2012), no item overlapped. The cluster analysis for EACOL's Form A demonstrated the same pattern as that for Form B.

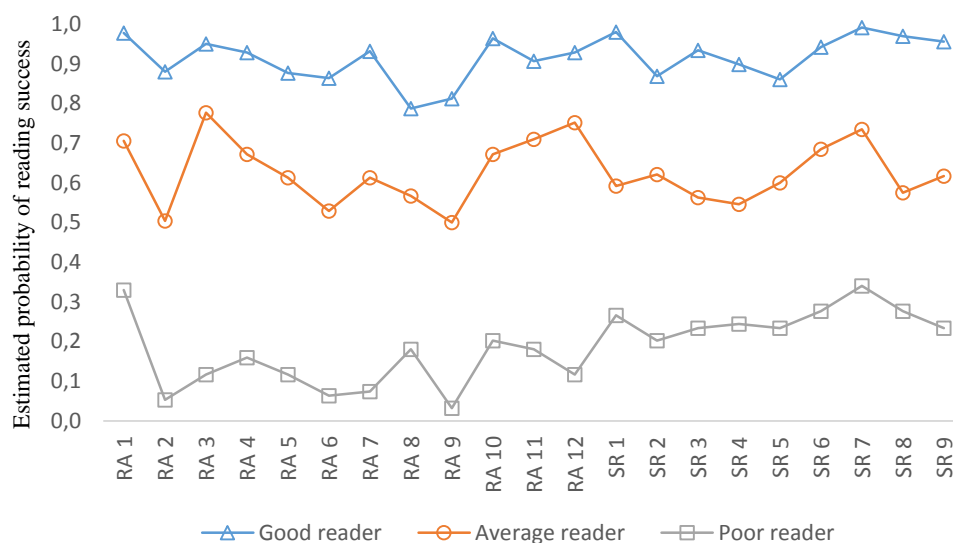


Figure 1. Two-step cluster analysis for Reading Aloud (RA) and Silent Reading (SR) items.

3.3 Descriptive analysis

Chart 1 shows the number of descriptors of good and poor readers in the final version of EACOL, excluding item 5 for Form A's Silent Reading. Form A is one third shorter than Form B because it has fewer descriptors of reading proficiency, since it is more lenient with the students it assesses, who are at the beginning of the literacy process.

Chart 1. EACOL number of items for each reader category and for the situations Reading Aloud (RA) and Silent Reading (SR)

Form	Subscale	Good readers	Poor readers	Total
A	RA	4	4	8
	SR	4	2	6
	Total	8	7	14
B	RA	6	6	12
	SR	5	4	9
	Total	11	10	21

No scale was eliminated due to internal inconsistency (opposing items answered more than twice) or incapability/difficulty of judgment of the teacher (four or more items answered “I do not know”). Although the alternative “I do not know” was chosen in just 1% of the possible cases, in 12% of the questionnaires there was at least one answer for this category. Thus, this alternative provided a relevant increment on the new scale. Another 1% of the scales returned with at least 1 item without answer; these items were scored with the same value as “I do not know.”

To verify the data distribution, skewness and kurtosis values were divided by the respective standard error, using a significance criterion of higher than 1.96 (Cramer & Howitt, 2004). All school grades demonstrated significant negative skewness: 2nd (-3.18), 3rd (-3.43), 4th (-5.35), and 5th grades (-5.70). A significant platykurtic distribution was found only in 4th (2.04) and 5th grades (2.03), thus showing a more uniform layout of data than the 2nd and 3rd grades. These statistical significances were confirmed using the Shapiro–Wilk normality test.

Standardization

Table 3 shows the norms for Forms A and B of the EACOL. The scores of the 4th and 5th grades did not differ numerically, and so these groups were combined.

Table 3. *Percentile norms and classification for raw EACOL scores by school grade*

Scales	School grade			Percentile	Classification
	2 nd	3 rd	4 th –5 th		
Reading Aloud	0–7	0–12	0–13	0–25%	Poor reader
	8–12	13–18	14–19	26–50%	Average reader –
	13–14	19–21	20–22	51–75%	Average reader +
	15–16	22–24	23–24	76–100%	Good reader
Silent Reading	0–6	0–9	0–11	0–25%	Poor reader
	7–10	10–14	12–15	26–50%	Average reader –
	11	15–18	16–18	51–75%	Average reader +
	12	-	-	76–100%	Good reader
Total score	0–15	0–23	0–26	0–25%	Poor reader
	16–23	24–32	27–34	26–50%	Average reader –
	24–26	33–39	35–39	51–75%	Average reader +
	27–28	40–42	40–42	76–100%	Good reader

4. Discussion

By assessing the EACOL in Brazil, the present study provides information that can be of use in developing an effective tool that is relevant to education policymakers, teachers, principals, parents, and pupils. Researchers, as external advisers, can play a pivotal role as catalysts for positive actions or informed reflections by these educational stakeholders. We hope the resubmission of the EACOL to this new validation and standardization study will stimulate teachers to carry out systematic evaluations of their students in elementary school, which, as the evidence shows, is an important way to prevent reading failure.

This screening instrument could be easily adapted to other countries, especially those that struggle with teaching Portuguese language, for instance, those with low amount of people aged 15 and over that can read and write: Guinea-Bissau (55.3%), Mozambique (56.1%), East Timor (58.3%), São Tomé and Príncipe (69.5%) and Angola (70.4%) (Central Intelligence Agency [CIA], 2014). In other nations of the Community of Portuguese-Speaking Countries, where literacy is above 90%, EACOL can be useful to screen children with risk of dyslexia; these places include Portugal and Cape Verde.

The new format of the EACOL significantly reduced the number of items in Form A (from 23 to 14) and Form B (from 27 to 21) without losing its validity. This should make the scale more attractive to the teacher, since it is now shorter and faster to complete.

Even with the new modifications, however, particularly with the addition of the answer “I do not know,” some scales were returned incomplete, reinforcing the conception that this problem may be due to some characteristic of the sample itself and not a failure of the scale. One theory is that the teachers in our sample prefer to decline to answer an item instead of admitting that they do not know about some aspect of their student’s reading performance. One way to minimize such behavior could be to add to the EACOL’s instructions the following

statement “Please always answer ‘I do not know’ in case of doubt; do not answer randomly or leave an item unanswered.”

On both Forms, the item correlations for the Reading Aloud subscale are stronger than those for Silent Reading. The descriptors for Reading Aloud, expressing explicit reading behavior, may be easier for the teacher to identify than those for Silent Reading. It can be inferred by the present results that the EACOL can be an effective instrument when teachers are instructed to select only students with poor reading ability.

Unlike Cogo-Moreira et al. (2012), this study found significant correlations between the EACOL, the CPM, and the SDQ. Cogo-Moreira et al. considered that the latter two measures would give EACOL good discriminant validity. Although the CPM is sometimes referred to as a non-verbal test, it requires language to process the information, and thus is better defined as a test of general cognitive ability (Vilhena, Pinheiro, & Gomes, 2014). Hence, a small-to-moderate positive correlation between the reading ability of the child and the CPM score is expected (Carver, 1990).

Concerning the child’s psychiatric characteristics, as assessed by the SDQ, a small but significant negative correlation is also expected. Maughan and Carroll (2006) note that disruptive behaviors impede reading progress and also the reverse: reading failure exacerbates risk for behavior problems. Thus, unlike Cogo-Moreira et al. (2012), we argue that although the variables measured by CPM and SDQ have distinct theoretical construct domains, they are not independent from each other (cf. Vilhena, Pinheiro, & Gomes, 2014).

As the correlations of the EACOL with general cognitive ability and psychiatric symptoms ranged from small to moderate, it is important to consider whether the teacher is taking these domains into account in her/his evaluations of children’s reading. One way to do so is to compare these correlations with those between CPM and SDQ within the General RC. First, as the correlations between the CPM and the General RC were smaller than those with

the EACOL (0.09 reduction in the value of r), we can argue that teachers can distinguish children's general cognitive ability on the basis of their reading ability. On the other hand, the SDQ had a bigger correlation with the EACOL than with the General RC (an additional 0.12 in the value of r). Although small, this correlation indicates that the teacher takes the child's behavior into consideration in his or her judgment.

As the scale was not designed to address children with excellent reading performance, an increase in the number of children in the "good" ability category occurred. This is demonstrated, for example, by the ceiling effect in Silent Reading on Form B and by the significant negative skewness distribution in all grades.

On the other hand, given the numerically wide range of scores, the EACOL is an effective scale to screen for poor readers, who should in any case be the first focus for early educational interventions in schools. The strong concordance between the reading task and the EACOL of children with poor ability is in agreement with the literature, which has shown that teachers are more accurate in the assessment of poor readers, identifying 89% of children with this type of performance (e.g., Capellini, Tonelotto, & Ciasca, 2004).

5. Conclusion

Reading ability is one of the most important competences in the modern world, essential to educational, professional, and social achievements. For this reason, it is of utmost relevance to create and/or adapt scientific validated instruments for early detection of poor reading skills and risk of dyslexia. With this purpose in mind, the EACOL was developed to be a quick and efficient instrument to guide educational stakeholders in assessing the Reading Aloud (speed and accuracy in word recognition, prosody and comprehension) and the Silent Reading (text comprehension and synthesis) of elementary-school children. Furthermore, this instrument can be adapted to other countries with Portuguese as the official language or to other orthographies.

6. Conflict of Interest

The authors declare no conflicts of interest.

7. Funding

This work was supported by the Pró-Reitoria de Pesquisa da Universidade Federal de Minas Gerais and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant No. 134357/2013-2). These bodies had no involvement in the study design; collection, analysis and interpretation of data; writing of the report; or decision to submit the article for publication.

8. References

- Angelini, A. L., Alves, I. C. B., Custódio, E. M., & Duarte, W. F. (1999). *Manual. Matrizes Progressivas Coloridas de Raven* [Manual: Raven's Coloured Progressive Matrices]. São Paulo: Casa do Psicólogo.
- Baker, S. F., & Ireland, J. L. (2007). The link between dyslexic traits, executive functioning, impulsivity and social self-esteem among an offender and non-offender sample. *International Journal of Law and Psychiatry*, 30, 492–503. DOI: 10.1016/j.ijlp.2007.09.010
- Capellini, S. A., Oliveira, A. M., & Cuetos, F. (2012). *PROLEC: Provas de avaliação dos processos de leitura* [Reading Processes Assessment Battery]. 2 ed. São Paulo: Casa do Psicólogo.
- Capellini, S. A., Tonelotto, J. M. F., & Ciasca, S. M. (2004). Learning performance measures: formal evaluation and teachers' opinion. *Revista Estudos de Psicologia, PUC-Campinas*, 21, 79–90. DOI: 10.1590/S0103-166X2004000200006
- Carreira-Perpiñán, M. Á. (1997). *A review of dimension reduction techniques*. Sheffield: Department of Computer Science, University of Sheffield. DOI: 10.1.1.212.994

- Carver, R. P. (1990). Intelligence and reading ability in grades 2–12. *Intelligence*, *14*(4), 449–455. DOI: 10.1016/S0160-2896(05)80014-5
- Central Intelligence Agency [CIA] (2014). *The world factbook*. Washington, DC: Author.
- Cogo-Moreira, H., Ploubidis, G., De Avila, C., Mari, J. & Pinheiro, A. M. V. (2012). EACOL (Scale of Evaluation of Reading Competency by the Teacher): Evidence of concurrent and discriminant validity. *Neuropsychiatric Diseases and Treatment*, *2012*, 443–454. DOI: 10.2147/NDT.S36196
- Cramer, D., & Howitt, D. (2004). *The Sage Dictionary of Statistics: A practical resource for students in the social sciences*. Thousand Oaks: Sage.
- Cury, C. R., & Golfeto, J. H. (2003). Strengths and Difficulties Questionnaire (SDQ): a study of school children in Ribeirão Preto. *Revista Brasileira de Psiquiatria.*, *25*(3), 139–145. DOI: 10.1590/S1516-44462003000300005
- Dobson, S. (2012). Current trends and future tendencies: Developing sustainable assessment cultures. *Policy Futures in Education*, *10*(4), 366–373. DOI: 10.2304/pfie.2012.10.4.366
- Emmerich, W., Enright, M. K., Rock, D. A., & Tucker, C. (1991). *The development, investigation, and evaluation of new item types for the GRE analytical measure*. Princeton: Educational Testing Service.
- Goodman R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581–586. DOI: 10.1111/j.1469-7610.1997.tb01545.x
- Goodman, R. (2005). Questionário de Capacidades e Dificuldades (SDQ-Por). Retrieved October 13, 2012, from www.sdqinfo.com.

- Gove, A., & Cvelich, P. (2011). Early reading: Igniting education for all. A report by the Early Grade Learning Community of Practice. Revised Edition. Research Triangle Park, NC: Research Triangle Institute.
- Instituto Paulo Montenegro (2011). *INAF Brasil 2011, Indicador de Alfabetismo Funcional. Principais resultados*. São Paulo: IBOPE Inteligência.
- Lúcio, P. S., & Pinheiro, A. M. V. (2013). Escala de Avaliação da Competência da Leitura pelo Professor (EACOL): Evidências de validade de critério. *Temas em Psicologia, 21*, 499–511. DOI: 10.9788/TP2013.2-15
- Maughan, B., & Carroll, J. M. (2006). Literacy and mental disorders. *Current Opinion in Psychiatry, 19*, 350–55. DOI: 10.1097/01.yco.0000228752.79990.41
- Pasquali, L., Wechsler, S., & Bensusan, E. (2002). Matrizes Progressivas do Raven Infantil: um estudo de validação para o Brasil. *Avaliação psicológica, 1*(2), 95–110.
- Pinheiro, A. M. V. (2007). Banco de palavras de baixa frequência de ocorrência, para crianças brasileiras da 1ª à 4ª série do Ensino Fundamental, classificadas em termos de estrutura silábica, número de letras e regularidade para leitura e para escrita. In I. Sim-Sim & F. L. Vianna, *Para avaliação do desempenho da leitura* (pp. 119–130). Lisbon, Portugal: Ministério da Educação.
- Pinheiro, A. M. V. (2013a). EACOL: Uma escala da avaliação da competência de leitura de alunos pelo Professor: Os primeiros estudos de validação. In F. L. Viana, R. Ramos, E. Coquet, & M. Martins (Eds.), *Leitura, Literatura Infantil e Ilustração: Investigação e Prática Docente*. Braga, Portugal: Centro de Investigação em Estudos da Criança, Universidade do Minho.
- Pinheiro, A. M. V. (2013b). *Prova de Leitura e de Escrita de palavras*. Relatório Técnico Final aprovado pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais—FAPEMIG (FAPEMIG). Número do processo: APQ-01914-09.

- Pinheiro, A. M. V., & Costa A. E. B. (2005). Escala de avaliação de competência em leitura pelo professor. *Anais do Encontro Mineiro de Avaliação Psicológica, Resumos*, 1, 27.
- Pinheiro, A. M. V., & Costa, A. E. B. (2012). A Construção e Validação da Escala de Avaliação de Competência Leitura pelo Professor—EACOL. *In Anais do VII 9º. Encontro Nacional, 7º. Internacional de Investigação em Leitura, Literatura Infantil e Ilustração*, Braga/Lisbon.
- Pinheiro, A. M. V., & Costa, A. E. B. (2015). EACOL—Escala de Avaliação da Competência em Leitura Pelo Professor: Construção por meio de Critérios e de Concordância entre Juízes. *Psicologia: Reflexão e Crítica*, 28(1), 1–10. DOI: 10.1590/1678-7153.201528109
- Saur, A. M., & Loureiro, S. R. (2012). Psychometric properties of the Strengths and Difficulties Questionnaire: A literature review. *Estudos de psicologia (Campinas)*, 29(4), 619–629. DOI: 10.1590/S0103-166X2012000400016
- Snowling, M. J., Duff, F., Petrou, A., Schiffeldrin, J., & Bailey, A. M. (2011). Identification of children at risk of dyslexia: The validity of teacher judgements using “Phonic Phases.” *Journal of Research in Reading*, 34, 157–170. DOI: 10.1111/j.1467-9817.2011.01492.x
- Todos pela Educação (2013). Instituto Paulo Montenegro/IBOPE, Fund. Cesgranrio, Inep. Retrieved November 20, 2013, from http://www.todospelaeducacao.org.br_
- UNESCO (2007). *Literacy Initiative For Empowerment 2006–2015. Vision and Strategy Paper, 3rd edition*. Hamburg: UNESCO Institute for Lifelong Learning.
- Verbič, S. (2012). Information value of multiple response questions. *Psihologija*, 45, 467–485. DOI: 10.2298/PSI1204467V
- Vilhena, D. A., & Pinheiro, A. M. V. (2014). Reading Performance Screening Test: Validation and standardization. Manuscript submitted for publication.

- Vilhena, D. A., Pinheiro, A. M. V., & Gomes, C. M. A. (2014). Predictors of reading ability: A structural equation modeling with demographic variables, language, cognitive and behavior assessments. In A. M. V. Pinheiro, & D. A. Vilhena. *Políticas públicas: Abrindo caminhos para a identificação precoce da dislexia e para as melhores práticas de alfabetização* (p. 77–77). Belo Horizonte: Faculdade de Filosofia e Ciências Humanas, Universidade Federal de Minas Gerais.
- Vilhena, D. A., Sucena, A., Castro, S. L., & Pinheiro, A. M. V. (2014). Reading Performance Screening Test: An adapted version of Lobrot's L3 test for Brazilian Portuguese. Manuscript submitted for publication.

Appendix. Items and scoring rubric for the EACOL (Scale of Evaluation of Reading Competence by the Teacher).

Form A (2nd grade) contains only the underlined sentences, while Form B (3rd–5th grade) contains both underlined and non-underlined sentences. Each item is followed by the possible responses: “True,” “False,” “Sometimes,” and “I do not know.”

Evaluation of Reading Aloud

Item	True	False	Some- times	I do not know
1. <u>Reads but cannot tell what was read, even when stimulated with questions.</u>	0	2	1	1
2. <u>Reads with intonation compatible with the punctuation marks, expressing emotions and feelings according to the text read. For example, gives an intonation of questioning in the whole sentence when there is a question mark in the text. Gives intonation of joy or surprise in the whole sentence when there is an exclamation mark.</u>	2	0	1	1
3. <u>Reads very slowly, without rhythm, spelling out each syllable; does not observe the punctuation marks.</u>	0	2	1	1
4. Reads by spelling out both “new” and “known” words.	0	2	1	1
5. <u>Does not take into account the intonation compatible with the punctuation marks, reading in a monotone manner.</u>	0	2	1	1
6. Reads “new” words correctly.	2	0	1	1
7. Reads clearly, without “stumbling” or “swallowing” syllables. Someone who hears can understand what is being read.	2	0	1	1
8. Frequently makes mistakes when reading “new” words.	0	2	1	1
9. <u>Quickly and correctly reads both “known” and infrequent words.</u>	2	0	1	1
10. <u>Seems to have understood what was read when asked about the text read.</u>	2	0	1	1
11. Has great difficulty in reading aloud.	0	2	1	1
12. <u>Reads with rhythm, neither too slowly nor too fast.</u>	2	0	1	1
x. <u>Reads with difficulty the “known” words.</u> *Item present only in Form A.	0	2	1	1

Evaluation of Silent Reading

Item	True	False	Some- times	I do not know
1. <u>Can identify characters, places, and main ideas after the first reading.</u>	2	0	1	1
2. Reads without pronouncing words or without moving the lips, only moving the eyes.	2	0	1	1
3. <u>Not able to orally summarize what was read.</u>	0	2	1	1
4. <u>Is able to identify the subject from the title and vice versa.</u>	2	0	1	1
5. Cannot read without movements of the lips or without pronouncing the words.	0	2	1	1
6. <u>Is able to choose a title for passages with no title or even give an alternate title for titled passages.</u>	2	0	1	1
7. Does not identify characters, places, or main ideas. *Corresponds to item 5 in Form A, which was excluded.	0	2	1	1
8. <u>Can orally summarize the text read.</u>	2	0	1	1
9. <u>Does not identify the subject from the title or vice versa.</u>	0	2	1	1

TESTE DE DESEMPENHO DE LEITURA

Nome: _____ Sexo: Masc. / Fem. Data de hoje: ____ / ____ / ____

Data de Nascimento: ____ / ____ / ____ Ano Escolar: 2º / 3º / 4º / 5º Nome do (a) Professor (a): _____

Jogo de Treino

1. A menina vestiu uma (rosa, pipa, roda, rua, roupa).
2. A estação fica no meio da (unidade, metade, cidade, grande, onde).
3. Todos os cachorros têm quatro (olhos, balas, pipas, patas, dedos).
4. Ele inclinou-se sobre o poço e caiu no (fundo, segundo, funil, futuro, furado).

1. Você poderia limpar a sala com uma (tesoura, manga, pente, cenoura, vassoura).
2. Eu adoraria ir à praia para tomar um banho de (maior, melhor, mar, par, formar).
3. Um cômodo onde se guarda livros chama-se (pesca, biblioteca, banheiro, salão, pasto).
4. Pegue a sacola e vá comprar (notas, crianças, palavras, laranjas, parcelas).
5. Ele espremeu a mão na porta e começou a chorar aos (gritos, ruídos, escritos, vidros, frios).
6. Na brincadeira, eu e meu amigo sujamos nossa (letra, mão, pão, não, ponte).
7. É primavera e os jardins estão floridos com (novas, roupas, casas, rosas, folhas).
8. Ele ligou o rádio para ouvir as (delícias, corretas, notícias, coloridas, tabelas).
9. Ele quebrou o prato e por isso se (abaniu, imaginou, cutucou, desmaiou, machucou).
10. O meu tio, depois de muito estudar, tornou-se um (jacaré, ninho, médico, senhor, comércio).
11. O meu irmão fez uma viagem à África e trouxe uma bela (vila, estátua, miragem, esquina, tempestade).
12. Minha mãe disse para não comer o bolo porque ainda está (valente, gostoso, bondoso, dente, quente).
13. As pessoas se assustaram: a locomotiva saiu dos (ouvidos, trilhos, astros, traços, troncos).
14. Quando for dormir, espero que tenha bons (sonhos, preços, cozidos, sorrisos, tecidos).
15. Vou lavar a louça amanhã de manhã porque estou com sono e prefiro ir para a (mata, pata, gata, cama, cesta).
16. Por que você não usa a faca para comer a (chave, chega, cheia, carne, cante)?
17. Da cratera do vulcão saem, de pouco em pouco, ondas de (selva, lava, cava, clava, mala).
18. Eles trabalham o dia inteiro, e à noite eles (conservam, expressam, processam, atravessam, descansam).

19. Quando a xingam e a castigam, ela fica (chateada, atrasada, deitada, empregada, cruel).
20. Um homem que dirige um veículo é chamado de (mecânico, companheiro, cientista, motorista, maquinista).
21. Ela saiu correndo de casa, por isso esqueceu sua (prosa, boba, bolsa, bolha, cebola).
22. Os nossos vizinhos compraram um cão grande e mau para ficar na porta da casa de (corda, girafa, nação, farda, guarda).
23. Nas noites de inverno as gotas de chuva são (tias, vazias, frias, salas, velas).
24. Se colocarmos o rádio muito alto, podemos incomodar os (viadutos, joelhos, partidos, vizinhos, passeios).
25. Nós fomos de carro até o parque, onde nos sentamos na grama para comer o nosso (lanche, plante, cheiro, rugido, ache).
26. Dentre todos os jogos, você prefere ping-pong, sinuca, dominó ou (portas, cartas, tortas, rins, fartas)?
27. O marido de uma filha é para a mãe dessa filha o (gênio, gentil, genro, generoso, general).
28. Aconteceu uma coisa engraçada a um pescador: ele pescou uma (lula, truta, carpa, sardinha, bota).
29. Nós fomos passear na praia e pegamos na areia algumas (tochas, conchas, colinas, manchas, colchas).
30. Todos saíram de casa para ver os estragos provocados pela (explosão, expansão, extinção, excursão, exceção).
31. As geladeiras evitam que a comida fique (enferrujada, estragada, desligada, resfriada, morta).
32. Já que está muito quente aqui, por que você não liga o (cobertor, colchão, ventilador, carregador, corredor)?
33. Quando andar na rua, é preciso ter muita atenção aos carros para não ser (enrolado, planejado, acabado, controlado, atropelado).
34. Eles combinaram de ir assistir à corrida no próximo domingo porquê gostam de ver os carros correrem na (pista, pasta, cesta, rota, blusa).
35. O mágico, ao pôr uma faca na palma da mão, nos deixou (contratados, sentados, entrevistados, assustados, devastados).
36. As pessoas gostam do que é novidade porque isso satisfaz a sua (curiosidade, dignidade, honestidade, vaidade, justiça).

ESCALA DE AVALIAÇÃO DA COMPETÊNCIA DE LEITURA PELO PROFESSOR (EACOL) – FORMA A

Aluno: _____ Nascimento: ____ / ____ / _____ Sexo: M / F Ano Escolar: 2º
 Professora: _____ Escola: _____ Data de Hoje: ____ / ____ / _____

Nº	AVALIAÇÃO DA LEITURA EM VOZ ALTA	Verdade	Falso	Às vezes	Não sei
1	Demonstra ter entendido o que leu quando indagado sobre o texto lido.				
2	Lê com entonação compatível com a pontuação, expressando emoções e sentimentos de acordo com o texto lido. Por exemplo: dá entonação de questionamento, em toda a sentença, quando há sinal de interrogação no texto; dá entonação de alegria ou de surpresa, em toda a sentença, quando há sinal de exclamação.				
3	Lê muito devagar, sem ritmo, soletrando cada sílaba, não observando a pontuação.				
4	Lê de forma rápida e correta as palavras “conhecidas”.				
5	Lê com ritmo, nem tão devagar, nem tão rápido.				
6	Lê com dificuldade as palavras “conhecidas”.				
7	Não observa a entonação compatível com os sinais de pontuação, fazendo uma leitura em um só tom.				
8	Lê, mas não sabe contar o que leu, nem quando estimulado com questões.				

Nº	AVALIAÇÃO DA LEITURA SILENCIOSA DE UM TEXTO	Verdade	Falso	Às vezes	Não sei
01	É capaz de identificar personagens, lugares e ideias principais após a primeira leitura.				
02	Não é capaz de resumir oralmente o que leu.				
03	É capaz de identificar o assunto a partir do título ou vice-versa.				
04	É capaz de escolher um título para passagens apresentadas sem título, ou mesmo um título alternativo para passagens com título.				
05	É capaz de resumir oralmente o que leu.				
06	Não identifica o assunto a partir do título ou vice-versa.				

ESCALA DE AVALIAÇÃO DA COMPETÊNCIA DE LEITURA PELO PROFESSOR (EACOL) – FORMA B

Aluno: _____ Nascimento: ____ / ____ / _____ Sexo: M / F Ano Escolar: 3º / 4º / 5º

Professora: _____ Escola: _____ Data de Hoje: ____ / ____ / _____

Nº	AVALIAÇÃO DA LEITURA EM VOZ ALTA	Verdade	Falso	Às vezes	Não sei
01	Lê, mas não sabe contar o que leu, nem quando estimulado com questões.				
02	Lê com entonação compatível com a pontuação, expressando emoções e sentimentos de acordo com o texto lido. Por exemplo: dá entonação de questionamento, em toda a sentença, quando há sinal de interrogação no texto; dá entonação de alegria ou de surpresa, em toda a sentença, quando há sinal de exclamação.				
03	Lê muito devagar, sem ritmo, soletrando cada sílaba, não observando a pontuação.				
04	Lê soletrando tanto palavras “novas” quanto as palavras “conhecidas”.				
05	Não observa a entonação compatível com os sinais de pontuação, fazendo uma leitura em um só tom.				
06	Lê corretamente as palavras “novas”.				
07	Lê de forma clara, sem “atropelar” ou “engolir” as sílabas. Quem ouve a leitura entende bem o que está sendo lido.				
08	Frequentemente comete erros ao ler palavras “novas”.				
09	Lê de forma rápida e correta as palavras “conhecidas” e as palavras “pouco conhecidas”.				
10	Demonstra ter entendido o que leu quando indagado sobre o texto lido.				
11	Tem grande dificuldade para ler em voz alta.				
12	Lê com ritmo, nem tão devagar, nem tão rápido.				

Nº	AVALIAÇÃO DA LEITURA SILENCIOSA DE UM TEXTO	Verdade	Falso	Às vezes	Não sei
01	É capaz de identificar personagens, lugares e ideias principais após a primeira leitura.				
02	Lê sem pronunciar as palavras ou sem movimentar os lábios, apenas movimentando os olhos.				
03	Não é capaz de resumir oralmente o que leu.				
04	É capaz de identificar o assunto a partir do título ou vice-versa.				
05	Não consegue ler sem fazer movimentos de lábios ou sem pronunciar as palavras.				
06	É capaz de escolher um título para passagens apresentadas sem título, ou mesmo um título alternativo para passagens com título.				
07	Não identifica personagens, lugares ou ideias principais.				
08	É capaz de resumir oralmente o que leu.				
09	Não identifica o assunto a partir do título ou vice-versa.				