Http://online.unisc.br/seer/index.php/signo
ISSN on-line: 0104-6578
Doi:

Recebido em 30 de Julho de 2021 Aceito em 12 de Novembro de 2021
Autor para contato: pinheiroamva@gmail.com
douglasvilhena@ufmg.br

Words and pseudowords recognition test: content and external validity

Teste de reconhecimento de palavras e pseudopalavras: validades de conteúdo e externa

Ângela Maria Vieira Pinheiro

Universidade Federal de Minas Gerais – UFMG – Minas Gerais - Brasil

Douglas de Araújo Vilhena

Universidade Federal de Minas Gerais – UFMG – Minas Gerais - Brasil

Abstract: The aim of this study is to provide the theoretical framework and psychometric evidences of content and external validity for the 'Word Recognition Test' (TRP) and 'Pseudoword Recognition Test' (TRPp). A total of 598 participants, 52% male, from the 2nd up to the 5th year of Elementary Education, 7 up to 11 years old, stratified in 102 classrooms of eight state schools. The TRP was generated from a database of words, analyzed via Item Response Theory, which considers Accuracy and Reaction Time, with medium to high discrimination index and three levels of difficulty (low, medium and high). The words were balanced with the psycholinguistic variables of: occurrence frequency (High vs. Low), regularity grapheme–phoneme (Semiregular vs. Irregular) and extension (Short vs. Long). The pseudowords have the same orthographic structure and number of letters as the reference word. The scores for both instruments are given terms of Accuracy (words percentage read correctly) and Accuracy Rate (words number read correctly per minute). Content validity was strengthened by the stimuli phonemic transcriptions and the lexicality significant effects (Words vs. Pseudowords), occurrence frequency, regularity and extension. External validity (convergent, discriminant and criterion) was verified. In conclusion, the TRP and the TRPp demonstrated satisfactory psychometric evidences of content and external validity.

Keywords: dyslexia; educational measurement; learning disabilities; validation study; psychometrics

Resumo: O estudo objetiva prover o referencial teórico e as fontes de evidências psicométricas de validade de conteúdo e validade externa para o Teste de Reconhecimento de Palavras (TRP) e o Teste de Reconhecimento de Pseudopalavras (TRPp). Participaram 598 alunos, 52% do sexo masculino, do 2º ao 5º ano do Ensino Fundamental, de 7 a 11 anos, estratificados em 102 salas de aula de oito escolas estaduais. O TRP foi gerado a partir de um banco palavras, analisado via Teoria de Resposta ao Item que, por sua vez, considerou a Acurácia e o Tempo de Reação, com índices de discriminação médio a alto e três níveis de dificuldade (baixa, média e alta). As palavras foram equilibradas com as variáveis psicolinguísticas de: frequência de ocorrência (Alta vs. Baixa), regularidade grafema–fonema (Semi-regulares vs. Irregulares) e extensão (Curto vs. Longo). As pseudopalavras possuem a mesma estrutura ortográfica e número de letras da palavra de referência. Os escores de ambos os instrumentos são dados em Acurácia (porcentagens de palavras lidas corretamente) e Taxa de Acurácia (número de palavras lidas corretamente por minuto). A validade de conteúdo foi fortalecida com as transcrições fonêmicas dos estímulos e com os efeitos de lexicalidade (Palavras vs. Pseudopalavras), frequência, regularidade e extensão significantes. A validade externa convergente, discriminante e de critério foi verificada. Concluiu-se que o TRP e o TRPp demonstraram evidências psicométricas de validade de conteúdo e validade externa satisfatórias.

Palavras-chave: Avaliação educacional; Deficiências da aprendizagem; Dislexia; Estudo de validação; Psicometria.

1. Introduction

The Word Recognition Test (*Teste de Reconhecimento de Palavras*, TRP) and the Pseudoword Recognition Test (*Teste de Reconhecimento de Pseudopalavras*, TRPp) were developed by Pinheiro (2013) to assess the reading ability from the 2nd to the 5th grades of Elementary School. Vilhena and Pinheiro (2022b) provided for both instruments evidences of reliability (internal consistency, split-half and parallel forms) and validity of the internal structure. The near-maximum internal consistency (*KR-20* = 0.98) was considered strong evidence that the TRP and TRPp items measure the same construct (*i.e.*, graphemic decoding).

Both the Accuracy variable (percentage of words read correctly) and the Accuracy Rate (number of words read correctly per minute) were able to distinguish readers by school grade (2nd < 3rd < 4th < 5th year) (Vilhena & Pinheiro, 2022b). The Accuracy Rate variable demonstrated a standard normal distribution curve, being symmetrical and mesokurtic, without overlapping the confidence intervals between the school grades. As the instruments showed satisfactory psychometric evidences to assess children from the 2nd to the 5th year of elementary school, the norms were presented for the TRP and TRPp based on a random and stratified sample, with adequate statistical size.

Although Vilhena and Pinheiro (2022b) demonstrated satisfactory psychometric evidences of reliability and validity of the internal structure for the TRP and TRPp, it is fundamental to provide the theoretical framework of the instruments based on the psycholinguistic characteristics recommended by the word recognition theory, such as frequency of occurrence, regularity of the grapheme–phoneme relationship and number of letters. Thus, the present study aims to provide the theoretical framework and psychometric evidences of content validity and external validity (convergent, discriminant, and criterion) for the TRP and the TRPp.

2. Theoretical framework

Most cognitive models of word recognition, both for proficient readers (Coltheart *et al.*, 2001; Perry *et al.*, 2007) or developing readers (Ehri, 2005), involve at least two components: (A) phonological processing, that uses grapheme–phoneme conversion to recognize low-frequency or unknown words and (B) orthographic processing, that uses the orthographic lexicon to recognize frequent words, especially irregular.

Regardless of how the relationship between phonological and orthographic processing is conceptualized by the different models, at an operational level, high-frequency words are used to assess orthographic processing and low-frequency words and pseudowords are used to assess phonological processing. This is the dominant view in which word recognition is considered according to the English language, that has opaque orthography.

However, authors such as Paulesu *et al.* (2000), Seymour *et al.* (2003), Dehaene *et al.* (2015), Pinheiro and Scliar-Cabral (2017, 2018), and Scliar-Cabral (2019) argue that in transparent orthographies (*e.g.* Italian, Spanish and Brazilian Portuguese) the access to the meaning of words occurs, with small exceptions, via the phonological mental lexicon (phonological route).

Scliar-Cabral (2019) explains that the skilled reader, through the decoding of a small number of graphemes, quickly evokes the phonological representation of the word in the phonological lexicon, which is used to access its meaning in the semantic memory. The access to this memory via the orthographic representation of words in the orthographic lexicon, unlike opaque orthographies, occurs only to resolve ambiguities and irregularities in the pronunciation of irregular and homophones words.

As for the processing of pseudowords – non-existent in the phonological and orthographic lexicon and, consequently, without basic meaning in semantic memory, but elaborated according to the principles of the alphabetic system of the written language of reference – there is more consensus in the literature. This processing occurs through the use of the phonological route, although some stimuli may be

susceptible to written lexical influence, as demonstrated in many languages in experiments that use the priming paradigm (*e.g.*, Campbell, 1983; Kay & Marcel, 1981). Thus, the graphemes that compose the pseudowords are converted into phonemes through decoding. As decoding operates in serial order, it requires the use of the phonological working memory (phonological loop) to maintain the phonological representation generated for subsequent articulatory output in reading aloud (and repetition, not discussed in this study) (Baddeley *et al.*, 1998).

The phonological processing for pseudowords is less accurate and takes more time than that for known words, thus generating the lexicality effect. However, at least for transparent orthographies, the interpretation of this effect differs from what occurs in opaque orthographies. In transparent orthographies, the decoding process for known words is so fast that the use of the phonological route is imperceptible. One of the types of experiment to prove processing in this route, cited by Scliar-Cabral (2008), are those that use words that may have a homophone no-homograph correlate (*e.g.*, *sela* and *cinto*). Scliar-Cabral explains that, although such words are known, it takes longer for the response to be emitted, as they are presented in competitive contexts and evidence the utilization of the phonological route. In the same line of reasoning, Dehaene (2012) observed that the persistence of the immediate reading and the whole word hypothesis (basis of global reading methods) may be due to the fact that the cognitive operations involved in word recognition are automatic, developed through parallel processing and occurs outside of our awareness.

Another difference between the orthographies refers to how reading competence is expressed in transparent orthographies. Unlike the opaque, reading accuracy often reaches its maximum after a few months of formal instruction (Cossu *et al.*, 1995; Landerl & Wimmer, 2008), bringing implications for the elaboration of word reading tasks, as will be shown.

In Portuguese, for example, due to the simplicity of the syllabic structure (Duncan *et al.*, 2013) and the transparency in the grapheme–phoneme correspondence (Pinheiro, 2011; Scliar-Cabral, 2003, 2019), the reading aloud tasks that measure only the accuracy are not difficult enough to represent the highest levels of reading ability (Lucio *et al.*, 2018) and do not pose a challenge even for beginning readers or those with reading difficulties (Pinheiro, 2011).

3. Method

3.1 Participants

All participants provided informed consent and the Research Ethics Committee of the *Universidade Federal de Minas Gerais* approved all study procedures (CAAE identification number: 17754514.6.0000.5149), which were conducted in full accordance with the Code of Ethics of the Declaration Helsinki (World Medical Association, 2008) for research involving human.

Participated in the current study 598 volunteer pupils, 52.3% male, aged 7 to 11 years (M = 9.6 ± 1.1 years), from the 2nd to 5th grade of elementary school, all native speakers of Brazilian Portuguese.

Eight state schools were randomly selected according to a list provided by the Regional Education Superintendence with all Public State Elementary Schools registered in the city of Belo Horizonte, Minas Gerais, Brazil. Schools were stratified in different regions of the city, with no refusal by the institutions to participate in the research. In each of the 102 classrooms, six participants were randomly selected (drawing from the attendance list) to respond to a cognitive assessment. Part of the teachers (n = 72) completed a behavior and a reading scale for each of the participate (n = 452), with the other teachers being considered dropouts for not having delivered the scales in time.

The children's general cognitive ability, assessed by a psychologist using Raven's Colored Progressive Matrices (Angelini *et al.*, 1999), averaged in the 76th percentile, classified as above average. Additional details about the participants and the cognitive assessment battery to which they were submitted, including the description of the Instruments and Procedures, can be found in different studies (Pinheiro

et al., 2017; Vilhena *et al.*, 2016; Vilhena & Pinheiro, 2016, 2020).

3.2 Instruments

The Word Recognition Test (TRP) (Pinheiro, 2013) assesses the reading aloud of isolated words, composed of 88 words (4 training and 84 test). The items were generated from a word database, applied to a representative sample of children and analyzed using the Item Response Theory, which considered not only the pupil's right or wrong answer but also the reaction time spent reading each word. The selected words had medium to high levels of discrimination and three levels of difficulty (low, medium, and high).

TRP items varied in terms of their frequency of occurrence, with 40 words being classified as high frequency [minimum 130 occurrences per million (*e.g.*, *animais*, *caderno*, *terra*)] and 44 low frequency [6 to 47 occurrences per million (*e.g.*, *atleta*, *monarca*, *sacola*)]. This classification was based on Pinheiro's (2015) Brazilian Portuguese word occurrence count database, composed of 1,774,164 words (tokens) contained in 124 books (*i.e.*, Portuguese, Mathematics, Social Studies, and Sciences), adopted by the Ministry of Education from the 1st to the 5th grade of Elementary School in all Brazilian states.

In addition to the frequency control, the TRP items were classified into two levels of regularity according to the decoding of each grapheme that constitutes the words [34 semi-regular (SR) and 50 irregular (IR)] and three levels of length [37 short words (four or five letters), 26 medium words (six letters) and 21 long words (seven or eight letters)]. The classification of the grapheme–phoneme relationship, originally proposed by Pinheiro (2003), was redone according to the 'Matrix of increasing complexity in the decoding of isolated words in Brazilian Portuguese' (Scliar-Cabral *et al.*, 2022). In this matrix, the words (Tables 1 and 2) are grouped into two categories: (1) Semi-regular – words with one or more than one grapheme Dependent on Graphemic Context [2DCG (*e.g.*, *nada*), 2DCG (*e.g.*, *fala*) and 5DCG (*e.g.*, *revista*)]; and (2) Irregular – words where the value of one grapheme is unpredictable, combined with one or more grapheme–phoneme correspondences DCG [IR+2DCG (*e.g.*, ***e***), IR+3DCG (*e.g.*, *co**m**e*), IR+4DCG (*e.g.*, *esp**e**ra*) and IR+5DCG (*e.g.*, *resp**o**sta*)].

The Pseudoword Recognition Test (TRPp) (Pinheiro, 2013) assesses the reading aloud of isolated pseudowords, composed of 88 items (4 training and 84 testing). The TRPp was built from the TRP, with each pseudoword maintaining the same orthographic structure and number of letters as the reference word (*e.g.*, *pederno*, *clandas*, *verra*) (Table 1 and Table 2).

The TRP and TRPp were presented on plastic cards, each containing 11 lines, with six to nine words per line, printed in black ink on A4 white paper, in Arial font, size 14. To provide evidence of reliability and to avoid the order effect, each instrument has three parallel forms, whose versions vary only in relation to the order of the words (Version A, B, and C) and of the pseudowords (Version D, E, and F).

Both TRP and TRPp have two types of scores: Accuracy (percentages of words read correctly) and Accuracy Rate (number of words read correctly per minute). For comparison between studies, the calculation of the variable Accuracy is given in percentage, since the reading instruments of isolated words present a difference in the number of items. The Accuracy variable is calculated by the total number of words or pseudowords read correctly (maximum of 84 points), divided by 84, times one hundred (*e.g.*, 75 words read correctly correspond to an Accuracy of 89%). The Accuracy Rate variable is the total number of words or pseudowords read correctly (maximum 84 points), times sixty, divided by the total time in seconds (*e.g.*, 75 words read correctly in 120 seconds correspond to an Accuracy Rate of 37.5 words per minute).

The Reading Test: Sentence Comprehension (TELCS) was used to assess sentence reading comprehension (Vilhena *et al.*, 2016; Vilhena & Pinheiro, 2020, 2022a). The TELCS is printed on both sides of an A4 sheet, composed of 40 isolated sentences (4 training and 36 testing), with the last word always being omitted. Five words are offered (multiple-

choice style), with only one alternative that makes sense to the sentence [*e.g.*, My uncle, after much study, became one (alligator, nest, doctor, sir, trade)]. TELCS must be answered within a maximum of five minutes. One point is scored for each correct answer and zero for incorrect or omitted ones (maximum: 36 points). The raw TELCS score corresponds to the number of sentences completed correctly within the time, with conversion to Percentile of reading performance rating. The TELCS demonstrated satisfactory psychometric properties, with sources of evidence of reliability (Cronbach's alpha, Spearman-Brown coefficient, test-retest), content validity (adaptation, psycholinguistic controls, item description), internal structure validity (effects of education and age, normal distribution), external validity (convergent, discriminant, criterion) and standardization (Machado & Maluf, 2019; Medina *et al.*, 2018; Medina & Guimarães, 2019; Pinheiro *et al.*, 2017; Vilhena & Pinheiro, 2016, 2020; Vilhena *et al.*, 2016).

The Reading Comprehension Test of the Reading Processes Assessment Tests (PROLEC-Text) (Capellini *et al.*, 2012) was used to assess the semantic processes. It consists of four short texts and four literal questions about each of them (maximum: 16 points), which requires extracting explicitly stated information in the text.

The Scale of evaluation of reading competence by the teacher (EACOL) is composed of operational descriptors of reading ability that can be recognized by the teacher (Pinheiro *et al.*, 2022; Pinheiro & Costa, 2015; Vilhena & Pinheiro, 2016). The reading aloud scale items measure speed and accuracy in word recognition, prosody, and comprehension; while the silent reading scale items measure comprehension and synthesis ability. Vilhena and Pinheiro found satisfactory psychometric indices for the following scoring criteria: a score of zero for predictors of reading difficulty, two points for predictors of good readers, and one point for predictors of fair readers.

Raven's Colored Progressive Matrices (CPM) (Angelini *et al.*, 1999) was used to measure the general cognitive ability of children aged between 5 and 11 years. CPM assesses analogical reasoning, which is the ability to infer relationships between objects or elements (Pasquali *et al.*, 2002). CPM consists of 36 items divided into three sets of 12 (A, Ab, B), organized according to increasing inter and intraset difficulty. The task is to complete a figure at the top of a page with one of six options printed below, which involves understanding that the images are characterized by their differences, similarities, identity, change, symmetry, and orientation.

The Strengths and Difficulties Questionnaire (SDQ) (Fleitlich-Bilyk *et al.*, 2000), used to provide discriminant validity, is a brief scale for screening externalizing behaviors of children aged 4 to 16 years, although it also assesses positive social behavior. The unilateral Brazilian version was used, without impact supplement, with scores for teachers. This instrument has 25 items divided into five scales: emotional symptoms (anxiety / mood), conduct problems (aggression / delinquency), hyperactivity / inattention, peer relationship problems (social problems / withdrawn), prosocial behavior (empathy / positive relationships). The SDQ has adequate reliability and validity indices in 21 countries, including Brazil (Saur & Loureiro, 2012).

3.3 Procedures

All instruments were administered on the same day, in two sessions, each lasting an average of 15 minutes. In the first session, groups of up to 10 children were collectively submitted to TELCS and CPM. In the second, each child was individually submitted to TRP and TRPp (administered in sequence but in random order), followed by PROLEC-Text. The instruments were administered by a psychologist and six undergraduate students in Psychology at the Federal University of Minas Gerais. The teachers answered the EACOL and the SDQ, for each participant, during one week.

Individually, participants were required to read the TRP and the TRPp aloud with the highest number of correct answers and as quickly as possible. The test

is easy to administer and to score. It starts with the four training items, followed by the 84 test items, read by the participant from left to right, top to bottom. In the TRP and TRPp correction sheets, the examiner recorded the reading accuracy errors dichotomously (correct and wrong) and the total reading time in seconds. The evaluation was stopped after 10 consecutive errors. Only fluent reading for each item was considered accurate. The autocorrection (item read incorrectly but immediately corrected) was considered a wrong answer.

3.4 Data analysis

Statistical analyzes were performed using the IBM SPSS Statistics software (version 21.0, Chicago, Illinois, USA). For the convergent and discriminant external validity, the Pearson *r* correlation coefficient (parametric data) and Spearman *ρ* (non-parametric data) were calculated, with values between 0 to 0.39 considered weak; between 0.40 to 0.69 moderate; and greater than 0.70 considered strong (Dancey & Reidy, 2005).

To establish the clinical significance of the differences, Cohen's *d* values between 0.30 and 0.49 were considered as moderate effect sizes; and greater than 0.50 considered strong (Cohen, 1988). The coefficients of the linear regression with a scatter plot were estimated (Figure 1). Statistical significance was set at $p < 0.05$.

4. Results

4.1 Evidence of content validity

As evidence of content validity, the phonemic transcripts of each word and pseudoword were provided (Table 1 and Table 2). The reading tests stimuli were balanced according to four psycholinguistic characteristics: lexicality (words and pseudowords), frequency of occurrence (high and low), length (short, medium, and long) and regularity (semiregular and irregular) (Table 1 and Table 2).

The psycholinguistic effect of lexicality was demonstrated by superior performance in reading real words compared to reading pseudowords. The analysis of variance for repeated measures, with months of age as a covariate, showed a strong significant difference between the TRP and the TRPp, both for the variable Accuracy ($F_{(1.594)} = 37.9$, $p < 0.001$, $d = 0.82$) and for the Accuracy Rate ($F_{(1.594)} = 203.7$, $p < 0.001$, $d = 1.14$). This lexicality effect occurs even when considering only low-frequency words, which were read more accurately than pseudowords ($F_{(1.594)} = 158.5$, $p < 0.001$, $d = 1.02$).

The frequency effect was demonstrated by the greater accuracy in reading frequent words, when compared to reading infrequent words ($F_{(1.594)} = 10.5$, $p = 0.0017$, $d = 0.70$).

Regarding the grapheme–phoneme regularity effect, semi-regular words were read more accurately than irregular words ($F_{(1.594)} = 9.8$, $p = 0.0028$, $d = 0.72$).

The extension effect was demonstrated by the greater accuracy in reading short stimuli, when compared to long, in the TRP ($F_{(1.594)} = 6.0$, $p = 0.0017$, $d = 0.64$) and in the TRPp ($F_{(2.594)} = 6.8$, $p = 0.011$, $d = 0.66$).

Table 1. Classification of TRP's words with high frequency of occurrence, phonemic transcription, frequency per million, classification of regularity for reading, length, and the corresponding pseudoword in TRPp.

Words				Length	Corresponding Pseudowords	
Item	Phonemic transcription	Frequency per million	Regularity		Item	Phonemic transcription
animais	/a.ni.'majS/	1213,8	5SR	L	atinais	/a.ti.'najS/
bola	/'bɔ.la/	453,3	IR+2SR	C	fola	/'fɔ.la/ ou /'fo.la/
caderno	/ka'dɛR.nU/	2977,6	IR+4SR	L	pederno	/pe'dɛR.nU/ ou /pe'deR.nU/
colegas	/ko.'lɛ.gaS/	365,9	IR+5SR	L	ropegas	/ro.'pɛ.gaS/ ou /ro.'pe.gaS/
coluna	/ko.'lu.na/	201,0	5SR	M	copuna	/ko.'pu.na/
corpo	/'koR.pU/	331,6	IR+3SR	C	lorpo	/'lɔR.pU/ ou /'loR.pU/
correta	/ko.'ʁe.ta/	365,9	IR+3SR	L	morreta	/mo.'ʁɛ.ta/ ou /mo.'ʁe.ta/
depois	/de.'pojS/	1287,4	3SR	M	pelois	/pe.'lojS/
dias	/di.aS/	764,2	3SR	C	bias	/bi.aS/
dois	/'dojS/	1239,2	2SR	C	cois	/'kojS/
dona	/'do.na/	460,9	3SR	C	tona	/'to.na/
duas	/'du.aS/	913,3	3SR	C	fuas	/'fu.aS/
enorme	/e.'nɔR.mI/	133,1	IR+3SR	M	emorne	/e.'nɔR.nI/ ou /e.'noR.nI/
escola	/eS.'kɔ.la/	807,1	IR+4SR	M	esgala	/eS.'ga.la/
fala	/'fa.la/	298,9	2SR	C	nala	/'na.la/
festa	/'fɛS.ta/	348,6	IR+2SR	C	fesda	/'fɛS.da/ ou /'feS.da/
figura	/fi.'gu.ɾa/	484,6	5SR	M	migora	/mi.'gɔ.ɾa/ ou /mi.'go.ɾa/
forte	/'fɔR.tI/	308,8	IR+2SR	C	dorte	/'dɔR.tI/ ou /'doR.tI/
frutas	/'fru.taS/	301,7	4SR	M	frulas	/'fru.laS/
letras	/'le.traS/	380,2	IR+3SR	M	petras	/'pɛ.tra/ ou /'pe.tra/
logo	/'lɔ.gU/ ou /'lo.gU/	463,3	IR+2SR	C	vogo	/'vɔ.gU/ ou /'vo.gU/
medo	/'me.dU/	359,1	IR+1SR	C	nedo	/'nɛ.dU/ ou /'ne.dU/
meio	/'mejo/	616,4	IR+1SR	C	teio	/'tɛjo/ ou /'tejo/
menina	/me.'ni.na/	608,2	2SR	M	mapina	/ma.'pi.na/
metros	/'mɛ.trUS/	408,1	IR+3SR	M	fetros	/'fɛ.trUS/ ou /'fe.trUS/
nada	/'na.da/	360,1	2SR	C	mada	/'ma.da/
noite	/'noj.tI/	520,6	IR+2SR	C	noife	/'nɔj.fI/ ou /'noj.fI/
novo	/'nɔ.vU/	251,8	IR+1SR	C	dovo	/'dɔ.vU/ ou /'do.vU/
palavra	/pa.'la.vɾa/	794,8	1SR	L	balacra	/ba.'la.kɾa/
papai	/pa.'paj/	476,6	3SR	C	pafai	/pa.'faj/
perto	/'pɛR.tU/	327,5	IR+2SR	C	verto	/'vɛR.tU/ ou /'veR.tU/
plantas	/'plã.taS/	649,6	4SR	L	clandas	/'klã.daS/
primeira	/pri.'mej.ɾa/	312,6	IR+5SR	L	prifeira	/pri.'fej.ɾa/
problema	/pɾo.'ble.ma/	306,7	5SR	L	croplema	/kɾo.'blɛ.ma/ ou /kɾo.'ble.ma/
resposta	/ʁeS'pɔS.ta/	924,3	IR+5SR	L	lesposta	/leS'pɔS.ta/ ou /leS'poS.ta/
sala	/'sa.la/	325,4	4SR	C	dala	/'da.la/
terra	/'tɛ.ra/	1179,3	IR+1SR	C	verra	/'vɛ.ra/ ou /'ve.ra/
veja	/'ve.ʒa/	961,6	IR+1SR	C	deja	/'dɛ.ʒa/ ou /'de.ʒa/
verde	/'veR.dI/	341,7	IR+2SR	C	serde	/'sɛR.dI/ ou /'seR.dI/
vida	/'vi.da/	604,2	2SR	C	mida	/'mi.da/

Legend – SR: semi-regular [the numeral indicates the number of graphemes dependent on the graphemic context; grapheme of interest is underlined]; IR: irregular (grapheme with unpredictable realization is in bold). C = short length (4 to 5 letters); M = medium (6 letters); L = Long (7 or 8 letters).
Note. Graphemes in Brazilian Portuguese can consist of one or two letters, as in the case of oral digraphs (*e.g.*, ss, sç, rr, ch, lh) and nasals (e.g., nh, am, an, en, in).

Table 2. Classification of TRP's words with low frequency of occurrence, phonemic transcription, frequency per million, classification of regularity for reading, length, and corresponding pseudoword in TRPp.

Words				Length	Corresponding Pseudowords	
Item	Phonemic transcription	Frequency per million	Regularity		Item	Phonemic transcription
atleta	/a.'tlɛ.ta/	17,8	IR+4SR	M	adleto	/a.'dlɛ.tU/ ou /a.'dle.tU/
bandeja	/bã.'de.ʒa/	15,8	4SR	L	mandeja	/mã.'dɛ.ʒa/ ou /mã.'de.ʒa/
bengala	/bɛ̃.'ga.la/	5,5	4SR	L	mengala	/mɛ̃.'ga.la/
briga	/'bri.ga/	40,3	4SR	C	criga	/'kri.ga/
caneca	/ka.'nɛ.ka/	15,2	IR+4SR	M	taneco	/ta.'nɛ.kU/ ou /ta.'ne.kU/
capela	/ka.'pɛ.la/	15,3	IR+3SR	M	mapela	/ma.'pɛ.la/ ou /ma.'pe.la/
cava	/'ka.va/	10,5	3SR	C	pava	/'pa.va/
cocada	/ko.'ka.da/	18,0	5SR	M	gocapa	/go.'ka.pa/
colina	/ko'li.na/	26,0	4SR	M	lolima	/lo'li.ma/
colo	/'kɔ.lU/	36,6	IR+3SR	C	folo	/'fɔ.lU/ ou /'fo.lU/
cometa	/ko.'me.ta/	13,6	1IR+3SR	M	bometa	/bo.'mɛ.ta/ ou /bo.'me.ta/
danada	/da.'na.da/	10,5	3SR	M	nanada	/na.'na.da/
delicada	/de.li.'ka.da/	16,4	6SR	L	relipada	/re.li.'pa.da/
espera	/eS.'pɛ.ɾa/	38,0	1IR+4SR	M	esvera	/eS.'vɛ.ɾa/ ou /eS.'ve.ɾa/
estojo	/eS.'to.ʒU/	29,5	4SR	M	esdojo	/eS.'dɔ.ʒU/ ou /eS.'do.ʒU/
favela	/fa.'vɛ.la/	20,4	IR+3SR	M	pavelo	/pa.'vɛ.lU/ ou /pa.'ve.lU/
forno	/'foR.nU/	17,0	IR+2SR	C	dorno	/'dɔR.nU/ ou /'doR.nU/
frota	/'frɔ.ta/	14,9	IR+2SR	C	crata	/'kra.ta/
gaveta	/ga.'ve.ta/	20,6	IR+3SR	M	soveta	/so.'vɛ.ta/ ou /so.'ve.ta/
limonada	/li.mo.'na.da/	8,3	5SR	L	timomada	/ti.mo.'ma.da/
luta	/'lu.ta/	46,7	3SR	C	cuta	/'ku.ta/
maleta	/ma.'le.ta/	16,0	IR+2SR	M	caleta	/ka.'lɛ.ta/ ou /ka.'le.ta/
marreco	/ma.'Rɛ.kU/	18,7	IR+3SR	L	tarreco	/ta.'ʁɛ.kU/ ou /ta.'ʁe.kU/
materno	/ma.'tɛR.nU/	12,9	IR+3SR	L	catermo	/ka.'tɛR.mU/ou /ka.'teR.mU/
moderna	/mo.'dɛR.na/	21,5	IR+3SR	L	coderna	/ko.'dɛR.na/ ou /ko.'deR.na/
monarca	/mo.'naR.ka/	5,6	5SR	L	bomarca	/bo.'maR.ka/
moto	/'mɔ.tU/	17,9	IR+1SR	C	poto	/'pɔ.tU/ ou /'po.tU/
neto	/'nɛ.tU/	33,0	IR+1SR	C	leto	/'lɛ.tU/ ou /'le.tU/
picada	/pi.'ka.da/	11,6	4SR	M	ricada	/ri.'ka.da/
pista	/'piS.ta/	38,8	2SR	C	quista	/'kiS.ta/
prego	/'prɛ.gU/	26,9	IR+3SR	C	frego	/'frɛ.gU/ ou /'fre.gU/
queda	/'kɛ.da/	24,1	2IR+2SR	C	beda	/'bɛ.da/ ou /'be.da/
rabanete	/Ra.ba.'ne.tI/	13,8	IR+3SR	L	cabamete	/ka.ba.'mɛ.tI/ ou /ka.ba.'me.tI/
redonda	/Re.'dõ.da/	36,3	4SR	L	medonta	/me.'dõ.ta/
resta	/'RɛS.ta/	18,3	IR+3SR	C	rista	/'ʁiS.ta/
revista	/Re.'viS.ta/	37,2	5SR	L	bevista	/be.'viS.ta/
sacola	/sa.'kɔ.la/	33,9	IR+5SR	M	tavola	/ta.'vɔ.la/ ou /ta.'vo.la/
sapeca	/sa.'pɛ.ka/	13,1	IR+4SR	M	fapeca	/fa.'pɛ.ka/ ou /fa.'pe.ka/
tapete	/ta.'pe.tI/	36,5	IR+2SR	M	mapete	/ma.'pɛ.tI/ ou /ma.'pe.tI/
tijolo	/ti.'ʒo.lU/	15,5	IR+4SR	M	vijalo	/vi.'ʒa.lU/
torta	/'tɔR.ta/	27,0	IR+2SR	C	rorta	/'rɔR.ta/ ou /'roR.ta/
trevo	/'tre.vU/	6,4	IR+2SR	C	drevo	/'drɛ.vU/ ou /'dre.vU/
uniforme	/u.ni.'fɔR.mI/	14,0	IR+4SR	L	unicorne	/u.ni.'kɔR.nI/ ou /u.ni.'koR.nI/
viola	/vi.'ɔ.la/	29,0	IR+3SR	C	siola	/si.'ɔ.la/ ou /si.'o.la/

Legend – semi-regular [the numeral indicates the number of graphemes dependent on the graphemic context; grapheme of interest is underlined]; IR: irregular (all words have only one unpredictable grapheme, except *queda* that has two, hence the number 2 in front of the acronym IR, grapheme with unpredictable realization is in bold); C = short length (4 to 5 letters); M = medium (6 letters); L = Long (7 or 8 letters).

4.2 Evidences of external validity

As evidence of convergent external validity, the variables Accuracy and Accuracy Rate of the TRP and the TRPp showed significant bivariate correlations ($p < 0.001$), from moderate to strong, with each other and with all the cognitive assessment battery variables of (a) reading ability: TELCS; PROLEC-Text; EACOL; (b) cognitive ability: CPM; and (c) demographic data (Table 3).

As for discriminant external validity, both variables of TRP and TRPp showed weak correlations with externalizing behaviors assessed by the SDQ (r = -0.36 to -0.11) (Table 3).

Based on the estimated coefficients of the regression line of the scatter plot (Figure 1), there was a positive linear relationship between Accuracy Rate of the TRPp as a function of the TRP ($r = 0.90$; $y = 0.46x + 5{,}86$; $R^2 = 0.6869$; $p < 0.001$). The determination coefficient (R^2) indicated the predictive capacity of the tested linear model, with 68.7% of the Accuracy Rate variance of the TRPp being explained by the TRP.

The strongest correlations were between the TELCS sentence comprehension with the TRP (r = 0.84) and TRPp (r = 0.79) Accuracy Rates (Table 3), as these variables involve the accuracy within a limited time to perform.

The TRP and TRPp were correlated with the response of 72 teachers, instrumented by the EACOL with operational criteria to assess the reading competence of 452 participants (Pinheiro & Costa, 2015; Vilhena & Pinheiro, 2016). The TRP and TRPp Accuracy variables obtained moderate correlations with the EACOL Form A for the 2nd year (r = 0.67 and 0.63) and with the Form B for the 3rd to 5th year (r = 0.48 and 0.52) (Table 3). Likewise, the TRP and TRPp Accuracy Rates showed moderate correlations with Form A (r = 0.66 and 0.65) and with Form B of EACOL (r = 0.59 and 0.55).

Evidence of external criterion validity will be provided if the Accuracy and Accuracy Rate variables are able to predict criteria established by the literature, in addition to the already demonstrated psycho-linguistic effects of lexicality, frequency, regularity and extension.

As a first criterion, the Accuracy and Accuracy Rate of the TRP and TRPp must be important variables to predict school performance. Part of the participants (n = 316) were classified by school performance in the Portuguese Language and Mathematics Classes, according to the Final Grade (sum of the grades of the four academic semesters), which ranged from Grade A (maximum grade) to Grade C (minimum grade found). In the TRP, participants with Final Grade A had higher scores than those with Final Grade B and C, both for the variable Accuracy ($F_{(3.312)} > 16.8$, $p < 0.001$), and for the Rate of Accuracy ($F_{(1.314)} > 10.5$, $p < 0.001$). The same pattern occurred with the TRPp (Final Grade A > B > C), for both classes, both for the variable Accuracy ($F_{(2.313)} > 30.2$, $p < 0.001$) and for the Accuracy Rate ($F_{(2.313)} > 14.2$, $p < 0.001$).

According to the second criterion, males and females must present equivalent reading performance (Vilhena & Pinheiro, 2020). Analysis of variance did not show a significant gender difference for the variable Accuracy and Accuracy Rate, both for the TRP and for the TRPp ($p > 0.15$).

The third evidence of criterion validity was verified when it was possible to find, by comparing the results of the TRP with the TRPp (Figure 1), participants with specific deficits. The performance profile of readers, regarding the contrast in word and pseudoword recognition, was classified into four groups:

(a) *competent performance*: good reading of words and pseudowords (n = 477, 80%);

(b) *mixed deficit*: poor reading of words and pseudowords (n = 92, 15%);

(c) *specific deficit in word processing*: poor word reading and good pseudoword reading (n = 16, 3%); and

(d) *specific deficit in the processing of pseudowords*: good reading of words and poor pseudowords reading (n = 13, 2%).

Table 3. Convergent and discriminant external validation: Spearman's Correlation (Accuracy) and Pearson's Correlation (Accuracy Rate) of TRP and TRPp with reading variables, general cognitive ability, social behavior and demographics.

Category	Variable		TRP		TRPp	
			Accuracy	Accuracy Rate	Accuracy	Accuracy Rate
Reading	TRP	Accuracy	1	.69**	.74**	.68**
		Accuracy Rate	.69**	1	.58**	.90**
	TRPp	Accuracy	.74**	.58**	1	.68**
		Accuracy Rate	.68**	.90**	.68**	1
	TELCS: Sentences		.64**	.84**	.57**	.79**
	PROLEC-Text		.43**	.46**	.34**	.42**
	EACOL: Form A, 2nd grade		.67**	.66**	.63**	.65**
	EACOL: Form B, 3rd to 5th grades		.48**	.59**	.52**	.55**
Cognition	CPM: general intelligence		.38**	.40**	.39**	.39**
Behavior	Pro social behavior		.18**	.19**	.18**	.17**
	Emotional Symptoms		-.26**	-.24**	-.26**	-.23**
	Conduct problems		-.26**	-.17**	-.26**	-.18**
	Hyperactivity / Inattention		-.36**	-.30**	-.36**	-.29**
	Problems with peers		-.17**	-.11*	-.17**	-.11*
	Total Score		-.36**	-.28**	-.36**	-.28**
Demographics	School grade		.32**	.53**	.24**	.51**
	Age (in months)		.34**	.55**	.25**	.48**
	Portuguese Language Class		.38**	.31**	.36**	.34**
	Mathematics Class		.38**	.23**	.37**	.27**

Note. *p < .01, **p < .001. TRP: Word Recognition Test; TRPp: Pseudoword Recognition Test; Accuracy: percentage of words read correctly; Accuracy Rate: number of words read correctly per minute; TELCS: Reading Test: Sentence Comprehension; PROLEC-Text: Text Comprehension Test of the Reading Process Assessment Tests; EACOL: Scale of evaluation of reading competence by the teacher; CPM: Raven's Colored Progressive Matrices; SDQ: Strengths and Difficulties Questionnaire; Class: final score given by the teacher. Source: own elaboration, with part published in Vilhena and Pinheiro, 2016; Pinheiro et al., 2017; Vilhena and Pinheiro, 2016, 2020.

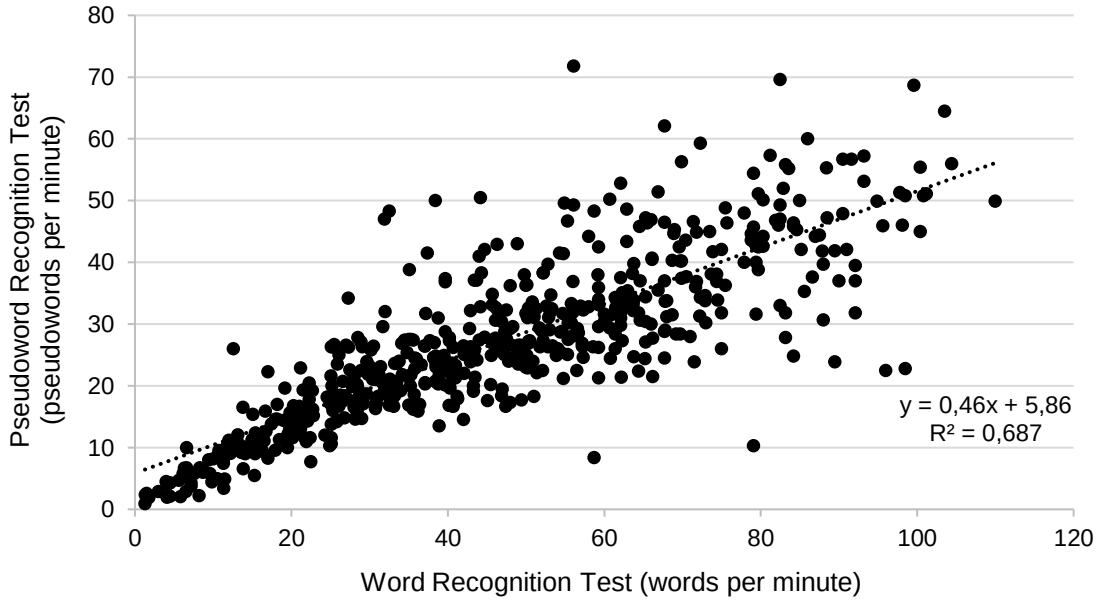

Figure 1. Scatter plot and linear regression of the Accuracy Rates of the Pseudoword Recognition Test (pseudowords per minute) as a function of the Word Recognition Test (word per minute).

The fourth evidence of criterion validity will be provided if the Accuracy values are consistent with the literature. As can be seen in Table 4, the mean Accuracy of the TRP (92%) and the TRPp (83%) presented values consistent with those found in different instruments (Capellini *et al.*, 2012; Justi & Justi, 2009; Lúcio *et al.*, 2012; Justi & Justi, 2009; Lúcio *et al.*, 2009; Salles *et al.*, 2013; Stivanin & Scheuer, 2007). No normative values were found for the Accuracy Rate variable.

Table 4. Percentage of Accuracy in the reading of words and pseudowords (Mean ± Standard Deviation) of the present study and of different reference instruments, organized in descending order of General Accuracy.

	Reference	Instrument	Number of items	School grade				
				Geral	2nd	3rd	4th	5th
Words (%)	Rodrigues *et al.* (2015)	TLPP	48	99 ± 2				99 ± 2
	Salles *et al.* (2013)	LPI	60	96	93	96	98	98
	Capellini *et al.* (2012)	PROLEC	30	96 ± 6	94 ± 11	95 ± 7	98 ± 4	98 ± 3
	Stivanin & Scheuer (2007)	PLPI	96	94		93 ± 9	93 ± 10	95 ± 9
	Present study	**TRP**	**84**	**92 ± 11**	**87 ± 16**	**92 ± 12**	**94 ± 7**	**96 ± 6**
	Justi & Justi (2009)	TLPI	40	92 ± 10			92 ± 10	
	Lúcio *et al.* (2009)	TDE	70	88 ± 1	81 ± 14	86 ± 10	90 ± 8	94 ± 7
	Lúcio *et al.* (2012)	TLVAPI	323	81	72	80	86	87
	Stein (1994)	TDE	70	80 ± 17	50 ± 39	84 ± 19	92 ± 7	95 ± 4
	Lúcio *et al.* (2018)	PLEP	48	77	64	76	83	85
			24	70	55	68	77	80
Pseudowords (%)	Salles *et al.* (2013)	LPI	20	93	93	93	93	95
	Capellini *et al.* (2012)	PROLEC	30	92 ± 9	91 ± 11	90 ± 10	95 ± 8	93 ± 8
	Justi & Justi (2009)	TLPpI	40	89 ± 10			89 ± 10	
	Rodrigues *et al.* (2015)	TLPP	24	88 ± 7				88 ± 7
	Present study	**TRPp**	**84**	**83 ± 14**	**77 ± 17**	**82 ± 15**	**82 ± 13**	**87 ± 11**

Note. LPI: Tarefa de Leitura de Palavras/Pseudopalavras Isoladas (Salles *et al.*, 2013); PLEP: Prova de Leitura e de Escrita de Palavras (Pinheiro, 2013); PLPI: Prova de leitura em voz alta de palavras isoladas (Stivanin, & Scheuer, 2007); PROLEC: Subteste de leitura de palavras e de pseudopalavras da PROLEC (Capellini *et al.*, 2012); TDE: Teste de Desempenho Escolar (Stein, 1994); TLPI: Teste de leitura de palavras isoladas (Justi, & Justi, 2009); TLPP: Tarefa de Leitura de Palavras e Pseudopalavras (Rodrigues *et al.*, 2015); TLPpI: Teste de leitura de pseudopalavras isoladas (Justi, & Justi, 2009); TLVAPI: Tarefa de leitura em voz alta de palavras isoladas (Pinheiro, 2004); TRP: Teste de Reconhecimento de Palavras (Pinheiro, 2013); TRPp: Teste de Reconhecimento de Pseudopalavras (Pinheiro, 2013).

5. Discussion

The present study aimed to provide the theoretical framework and psychometric evidences of content validity and external validity for the TRP and the TRPp. Both instruments presented satisfactory psychometric evidence to assess the graphemic decoding of pupils from the 2nd to the 5th grade of Elementary School. These results complement the reliability and internal structure validity found in previous studies (Cogo-Moreira *et al.*, 2012; Pinheiro, 2013; Pinheiro *et al.*, 2017; Vilhena, 2015; Vilhena *et al.*, 2016; Vilhena & Pinheiro, 2016, 2020, 2022b).

For content validity, the phonemic transcription of all items and a detailed description of the operational and constitutive definitions of the TRP and TRPp items were presented (Table 1 and Table 2). To test reading strategies, the assessment must include a list of stimuli that first considers the frequency level (balance of low and high-frequency words), then regular and irregular words, and finally, the balance of short and long stimuli. The number of items per test (N = 84) was sufficient to identify all four psycholinguistic effects: lexicality, frequency, regularity, and extension.

The lexicality effect was investigated by contrasting the reading of words with pseudowords. On average, participants read words half a second faster than the pseudowords. Both TRP variables had higher scores than the TRPp, significantly and with a strong effect size. This demonstrates that lexicality is a factor that facilitates Accuracy and Accuracy Rate, with increased precision and reduced reading time. The reading of pseudowords, as it is performed serially and without the support of the phonological lexicon, is more imprecise and time-consuming than the reading of words.

The explanation of the lexicality effect depends on the theoretical framework adopted. For the predominant view, represented by the dual-route model (Coltheart *et al.*, 2001), word reading offers an advantage over pseudoword reading because the first stimuli have previously stored representations in the orthographic lexicon.

Alternatively, in transparent orthographies, this advantage occurs because the words have previously stored representations not in the orthographic lexicon, but in the phonological lexicon. The decoding of all the graphemes of a given word allows the reader, who has the phonological representation that he 'hears' in his inner speech, to match it with the phonological word in his phonological lexicon, if it is a known word, to then look for it's basic meaning in semantic memory. The pronunciation of words that contain graphemes with unpredictable values must be memorized (Scliar-Cabral, 2019). Finally, the automatic and parallel form of lexical processing, regardless of its *locus* (orthographic or phonological lexicon), in contrast to non-lexical processing, confers an advantage for word reading over pseudoword reading.

For the TRP, the effects of frequency of occurrence, regularity, and length were also found, which, respectively, showed an advantage in terms of accuracy for frequent, regular, and short words compared to infrequent, irregular, and long words. As for the theoretical framework that explains the reading processes in transparent orthographies, the processing of real, frequent, regular, and short words facilitates the process of grapheme–phoneme conversion and subsequent access to the phonological representation of the word. Frequent words have their phonological representation better established in the phonological lexicon. Regular words elicit fewer errors and less processing time than irregular words, because the latter have unpredictable grapheme–phoneme relationships, which require decision-making about how they are pronounced. On the other hand, short words, in contrast to long words, require a smaller number of graphemes to be decoded so that the reader can access the phonological representation of the word. Finally, pseudowords do not have their pronunciations recorded in the phonological lexicon.

The absence of words in the Regular category was observed. According to Scliar-Cabral (personal communication), in written Brazilian Portuguese, words rarely have all graphemes decoded regardless of the graphemic context, because the vowel graphemes that are independent of this context are rare and there are no words in written Brazilian Portuguese without vowel graphemes.

The unpredictability is characterized when there are no rules that determine the value, that is, the phoneme into which the grapheme is converted. For example, for the graphemes <e> and <o> when they appear in paroxytone syllables there are two alternatives for decoding each one: <e> /e/ or /ɛ/; <o> /o/ or /ɔ/). According to Scliar-Cabral (2003), in Brazilian Portuguese there are still two types of unpredictability: (1) the value of the grapheme <x> in an intervocalic position (*e.g.*, *fixo*, *táxi*, *nexo*), except for the case where <e> is at the beginning of the word and <x> at the beginning of the syllable, as in *exato*, *exemplo* (there is no word in TRP with <x>); and (2) the value of the graphemes <gu>, <qu>, followed by <e> or <i>, converted, respectively, into /g/, /k/; or the value of the graphemes <g>, <u>, <q>, <u>, followed by <e> or <i>, converted, respectively, into /g/, /w/; /k/, /w/. With the new Portuguese orthographic agreement, these sequences of letters can perform respectively the graphemes <gu>, <qu>, representing the phonemes /g/, /k/ as in *guerra* and *quilo* or as the graphemes <g>, <u >, <q>, <u>, representing the phonemes /g/, /w/, /k/, /w/, as in *aguenta* and *cinquenta*. There is only one word in TRP in this category: *queda*.

Validity evidence based on relationships with external variables includes convergent, discriminant, and criterion validations. For convergent external validity, as expected, TRP and TRPP showed strong correlations with each other, both for Accuracy (ρ = 0.80) and for the Accuracy Rate (r = 0.89), which demonstrates a degree of linear statistical dependence between the test variables, as they present shared cognitive substrates. Likewise, TRP and TRPp showed significant ($p < 0.001$) moderate to strong correlations with different comparison reading instruments (see Table 3). These results complement those found for the TRP and the TRPp in a sample from São Paulo, where the instruments showed a high correlation with each other ($r = 0.92$, $p < 0.001$) and with the accuracy of text reading (r = 0, 92 and 0.87, $p < 0.001$) (Cogo-Moreira *et al.*, 2012).

The TRP and TRPp Accuracy Rate showed a strong correlation with the Reading Test: Sentence Comprehension (r = 0.84 and 0.79). Weak to moderate correlations were found between Accuracy (TRP and TRPp) and text comprehension assessed using literal questions of the PROLEC-Text (r = 0.43 and 0.34) (Pinheiro *et al.*, 2017). The moderate correlations between the Accuracy Rate (TRP and TRPp) and the PROLEC-Text (r = 0.46 and 0.42) are consistent with the weak to moderate correlations found between the variable 'syllables per minute' and 'understanding of text' (r = 0.30 to 0.43) in a sample of 188 Italian children from 4th and 5th grade (Bigozzi *et al.*, 2017). Weak correlations (r = -0.36 to -0.11) between Accuracy and Accuracy Rate and externalizing behaviors, investigated via SDQ, provide evidence of discriminant validity for TRP and TRPp, showing that the tests assess distinct constructs.

Four different evidences of criterion validity were provided for the TRP and the TRPp, demonstrating that the instruments are adequate to predict reading and school performance. First, TRP and TRPp are good predictors of the Final Grade received by pupils at the end of the school year in the Portuguese Language and Mathematics Classes. This result reveals that the degree of reading efficiency is reflected in school classes, with participants with poor academic performance (Final Grade B and C) presenting, on average, difficulty in Accuracy and Accuracy Rate.

As the second evidence of criterion validity, both the Accuracy variable and the Accuracy Rate of the TRP and the TRPp did not show a significant sex difference, which corroborates the literature (Athayde *et al.*, 2014; Lúcio *et al.*, 2018; Vilhena & Pinheiro, 2020).

The third evidence of criterion external validity reveals the importance of using the TRP and the TRPp together to verify specific deficits in reading performance. It was possible to verify four groups: (a) competent performance (80%); (b) mixed deficit (15%); (c) specific deficit in word processing (3%); and (d) specific deficit in pseudoword processing (2%).

When basing the clinical diagnosis of developmental dyslexia, it is possible to interpret the results according to different theoretical framework. For the dual-route model (Coltheart *et al.*, 2001),

participants with a specific deficit in word processing (3%) demonstrate limitations in spelling processing, referred to as 'lexical developmental dyslexia' or 'surface dyslexia'. On the other hand, participants with poor performance only in reading pseudowords (2%) have a specific deficit in phonological processing, referred to as 'phonological development dyslexia'.

When interpreting the results according to the theoretical framework that considers transparent orthographies, participants with a specific deficit in word processing demonstrate a deficit in accessing the phonological lexicon. On the other hand, participants with poor performance only in reading pseudowords have a specific deficit in phonological processing, maintaining access to representations stored in the phonological lexicon.

As the fourth criterion, the high mean Accuracy of the TRP (92%) is consistent with the literature (Capellini *et al.*, 2012; Justi & Justi, 2009; Lúcio *et al.*, 2009; Salles *et al.*, 2013; Stivanin & Scheuer, 2007). In general, the single-word reading tests are easy to read, with typical children reaching high levels of accuracy, consistent with the orthographic transparency for reading in Brazilian Portuguese. Most of the word reading assessment instruments presented an average of Accuracy above 85% as of the 2nd year of Elementary School (Table 4). Regarding the Accuracy of the TRPp, which progressed from 77% to 87% between school years (average = 83%), it was found that the selected pseudowords are, on average, more difficult to read than the other reference studies, whose averages ranged from 88 to 95%.

Lucio *et al.* (2018) showed that the Word Reading and Writing Test (PLEP) (Pinheiro, 2013) is the most difficult instrument to read, since both the 48-word and 24-word versions had the lowest Accuracy averages (70%) among the studies analyzed (Table 4). In addition to the fact that the PLEP is composed only of low-frequency words, it is possible that this result was due to the regularity classification of words, designed to test both reading and writing. Thus, PLEP words are semi-regular (and irregular) in both grapheme–phoneme and phoneme–grapheme correspondence. The rationale for the elaboration of stimuli with such a bidirectional classification of regularity is found in Pinheiro (2011, 2013).

Conclusion

The Word Recognition Test and the Pseudoword Recognition Test demonstrated satisfactory psychometric evidences of content validity and external validity for the assessment of reading ability in children from the 2nd to the 5th grade of Elementary Schools in Brazil, complementing the evidence of reliability and internal structure validity found in other studies. The theoretical framework and empirical analyses presented in the current article, and in previous studies, support that both TRP and TRPp can be considered as gold standard tests for assessing graphemic decoding.

References

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*. CETEPP.

Athayde, M. L., Giacomoni, C. H., Zanon, C., & Stein, L. M. (2014). Evidências de validade do subteste de leitura do teste de desempenho escolar. *Revista Psicologia: Teoria e Prática*, *16*(2), 131-140. http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-36872014000200012&lng=pt&tlng=pt

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review*, *105*(1), 158–173. http://dx.doi.org/10.1037/0033-295x.105.1.158

Bigozzi, L., Tarchi, C., Vagnoli, L., Valente, E., & Pinto, G. (2017). Reading Fluency as a Predictor of School Outcomes across Grades 4-9. *Frontiers in psychology*, *8*, 200. http://dx.doi.org/10.3389/fpsyg.2017.00200

Campbell, R. (1983). Writing nonwords to dictation. *Brain and language*, *19*(1), 153–178. http://dx.doi.org/10.1016/0093-934x(83)90061-5

Capellini, S. A., Oliveira, A. M., & Cuetos, F. (2012). *PROLEC: Provas de Avaliação dos Processos de Leitura* (2. ed.). Casa do Psicólogo.

Cogo-Moreira, H., Ploubidis, G., De Avila, C., Mari, J., & Pinheiro, Â. M. V. (2012). EACOL (Scale of Evaluation of Reading Competency by the Teacher): Evidence of concurrent and discriminant validity. *Neuropsychiatric Diseases and Treatment*, *8*(1), 443–454. https://doi.org/10.2147/NDT.S36196

Cohen, J. W. (1988). *Statistical power analysis for the behavioural sciences* (2. ed.). Lawrence Erlbaum Associates.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, *108*(1), 204–256. http://dx.doi.org/10.1037/0033-295x.108.1.204

Cossu, G., Giuliotta, M., & Marshall, J. C. (1995). Acquisition of reading and written spelling in transparent orthography. Two non-paralle processes? *Reading and Writing*, *7*, 9–22. https://doi.org/10.1007/BF01026945

Dancey, C., & Reidy, J. (2006). *Estatística Sem Matemática para Psicologia*: Usando SPSS para Windows. Artmed.

Dehaene, S. (2012). *Os neurônios da leitura - como a ciência explica a nossa capacidade de ler* (L. Scliar-Cabral, trad.). Penso.

Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, *16*(4), 234–244. https://doi.org/10.1038/nrn3924

Duncan, L. G., Castro, S. L., Defior, S., Seymour, P. H., Baillie, S., Leybaert, J., & Serrano, F. (2013). Phonological development in relation to native language and literacy: variations on a theme in six alphabetic orthographies. *Cognition*, *127*(3), 398-419. http://dx.doi.org/10.1016/j.cognition.2013.02.009

Ehri, L. C. (2005). Development of Sight Word Reading: Phases and Findings. In M. J. Snowling & C. Hulme (Eds.), *Blackwell handbooks of developmental psychology.* The science of reading: A handbook (pp. 135–154). Blackwell. https://doi.org/10.1002/9780470757642.ch8

Fleitlich-Bilyk, B., Cortázar, P. G., & Goodman, R. (2000). Questionário de Capacidades e Dificuldades (SDQ) [Strengths and Difficulties Questionnaire (SDQ)]. *Infanto: Revista de Neuropsiquiatria da Infância e Adolescência*, *8*(1), 44-50.

Justi, C. N. G., & Justi, F. R. R. (2009). Os efeitos de lexicalidade, freqüência e regularidade na leitura de crianças falantes do português brasileiro. *Psicologia: Reflexão e Crítica*, *22*(2), 163-172. https://dx.doi.org/10.1590/S0102-79722009000200001

Kay, J., & Marcel, T. (1981). One process not two in reading aloud: lexical analogies do the work of non-lexical rules. *Quarterly Journal of Experimental Psychology*, *33*(4), 397-413. https://doi.org/10.1080/14640748108400800

Landerl, K., & Wimmer, H. (2008). Development of Word Reading Fluency and Spelling in a Consistent Orthography: An 8-Year Follow-Up. *Journal of Educational Psychology*, *100*(1), 150-161. https://doi.org/10.1037/0022-0663.100.1.150

Lúcio, P. S., Pinheiro, Â. M. V., & Nascimento, E. (2009). O impacto da mudança no critério de acerto na distribuição dos escores do subteste de leitura do Teste de Desempenho Escolar. *Psicologia em*

Estudo, *14*(3), 593-601. http://dx.doi.org/10.1590/S1413-73722009000300021

Lúcio, P. S., Moreira, H. C., Kida, A. S. B., Carvalho, C. A. F., Pinheiro, Â. M., Mari, J. J., & Avila, C. R. B. (2018). Word Decoding Task: Item Analysis by IRT and Within-Group Norms. *Psicologia: Teoria e Pesquisa*, *34*, e3437. https://dx.doi.org/10.1590/0102.3772e3437

Lúcio, P. S., Moura, R. J., Nascimento, E., & Pinheiro, Â. M. V. (2012). Construção de uma tarefa de leitura em voz alta de palavras: análise psicométrica dos itens. *Psicologia: Reflexão e Crítica*, *25*(4), 662-670. http://dx.doi.org/10.1590/S0102-79722012000400005

Machado, M. S. M., & Maluf, M. R. (2019). Como evolui a compreensão da leitura em alunos do ensino fundamental. *Psicologia da Educação*, *49*, 57-66. http://dx.doi.org/10.5935/2175-3520.20190019

Medina, G. B. K., & Guimarães, S. R. K. (2019). Reading of Students with Developmental Dyslexia: Impacts of an Intervention with Phonic Method Associated with the Executive Functions Stimuli. *Revista Brasileira de Educação Especial*, *25*(1), 155–174. http://dx.doi.org/10.1590/s1413-65382519000100010

Medina, G. B. K., Souza, F. F., & Guimarães, S. R. K. (2018). Funções executivas e leitura em crianças brasileiras com dislexia do desenvolvimento. *Revista Psicopedagogia*, *35*(107), 168–179. http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0103-84862018000200005&lng=pt&tlng=pt

Pasquali, L., Wechsler, S. M., & Bensusan, E. (2002). Matrizes Progressivas do Raven Infantil: um estudo de validação para o Brasil [Raven's Colored Progressive Matrices for Children: a validation study for Brazil]. *Avaliação Psicológica*, *1*(2), 95–110. http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712002000200003&lng=pt&tlng=pt.

Paulesu, E. *et al.* (2000). A cultural effect on brain function. *Nat Neurosci*, *3*(1), 91-96. https://doi.org/10.1038/71163

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273-315. http://dx.doi.org/10.1037/0033-295X.114.2.273

Pinheiro, Â. M. V. (2004). *Banco de palavras de alta e 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*. Relatório Técnico Final aprovado pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG. Número do processo: DC/SOT-1202/2004.

Pinheiro, Â. 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, R. Mousinho & S. A. Capellini (Eds.). *Dislexia: Novos temas, novas perspectivas* (pp. 131–146). Wak.

Pinheiro, Â. M. V. (2013). *Provas de Leitura e de Escrita de palavras e de pseudopalavras*. Relatório Técnico Final aprovado pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG. Número do processo: APQ-01914-09.

Pinheiro, Â. M. V. (2015). Frequency of Occurrence of Words in Textbooks Exposed to Brazilian children in the Early Years of Elementary School. Updated database from 1996. CHILDES – Child Language Data Exchange System. http://childes.talkbank.org/derived

Pinheiro, Â. 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. http://dx.doi.org/10.1590/1678-7153.201528109

Pinheiro, Â. M. V., & Scliar-Cabral, L. (2018). Dislexia: Causas e Consequências. Editora UFMG e-book. https://www.ufmg.br/ead/epubs/index.php/2018/09/17/dislexia-causas-e-consequencias/

Pinheiro, Â. M. V., & Scliar-Cabral, L. (2017). *Dislexia: Causas e Consequências.* Editora UFMG.

Pinheiro, Â. M. V., Costa, A. E. B., & Vilhena, D. A. (2022). *Escala de Avaliação da Competência em Leitura Pelo Professor – EACOL: Manual Técnico*. NILAPRESS Editora.

Pinheiro, Â. M. V., Vilhena, D. A., & Santos, M. A. C. (2017). PROLEC – Provas de Avaliação dos Processos de Leitura: análise de suas características psicométricas. *Trends in Psychology* [*Temas em Psicologia*], *25*(3), 1067-80. http://dx.doi.org/10.9788/tp2017.3-08

Rodrigues, J. C., Nobre, A. P., Gauer, G., & Salles, J. F. (2015). Construção da tarefa de leitura de palavras e pseudopalavras (TLPP) e desempenho de leitores proficientes. *Temas em Psicologia*, *23*(2), 413-29. http://dx.doi.org/10.9788/TP2015.2-13

Salles, J. F., Piccolo, L. R., Zamo, R. S., & Toazza, R. (2013). Normas de desempenho em tarefa de leitura de palavras/pseudopalavras isoladas (LPI) para crianças de 1º ano a 7º ano. *Estudos e Pesquisas em Psicologia*, *13*(2), 397-419. https://doi.org/10.12957/epp.2013.8416

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. https://doi.org/10.1590/S0103-166X2012000400016

Scliar-Cabral, L. (2003). *Princípios do Sistema Alfabético do Português do Brasil*. Editora Contexto.

Scliar-Cabral, L. (2008). Processamento bottom-up na leitura. Veredas on-line – *Psicolinguística*, *2*(1), 24-33. https://www.ufjf.br/revistaveredas/files/2009/12/artigo02.pdf

Scliar-Cabral, L. (2013). *Sistema Scliar de alfabetização – Fundamentos*. Editora Lili.

Scliar-Cabral, L. (2019). Precisamos de mais de 50.000 palavras no léxico ortográfico para ler? *Letras de Hoje*, *54*(2), 122-131. http://dx.doi.org/10.15448/1984-7726.2019.2.34510

Scliar-Cabral, L., Martins, P. S. R., & Pinheiro, Â. M. V. (2022). Matriz da complexidade crescente na decodificação e decodificação de palavras isoladas no português brasileiro. Em preparação.

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

Stein, L. (1994). *TDE: Teste de desempenho escolar: Manual para aplicação e interpretação*. Casa do Psicólogo.

Stivanin, L., & Scheuer, C. I. (2007). Tempo de latência para a leitura: influência da freqüência da palavra escrita e da escolarização. *Revista da Sociedade Brasileira de Fonoaudiologia*, *12*(3), 206-13. http://dx.doi.org/10.1590/S1516-80342007000300008

Vilhena, D. A., & Pinheiro, A. M. V. (2016). Revised version of the Scale of Evaluation of Reading Competence by the Teacher: final validation and standardization. *Universitas Psychologica*, *15*(4), 1-13. http://dx.doi.org/10.11144/Javeriana.upsy15-4.efvs

Vilhena, D. A., & Pinheiro, Â. M. V. (2020). Reliability, Validity and Standardization of the Reading Test: Sentence Comprehension [Fidedignidade, Validade e Normatização do Teste de Leitura: Compreensão de Sentenças]. *Psicologia: Teoria e Pesquisa*, *36*, e36325. https://dx.doi.org/10.1590/0102.3772e36325

Vilhena, D. A., & Pinheiro, Â. M. V. (2022a). *Teste de Leitura: Compreensão de Sentenças – TELCS: Manual Técnico*. NILAPRESS Editora.

Vilhena, D. A., & Pinheiro, Â. M. V. (2022b). Fidedignidade, Validade Interna e Normatização dos Testes de Reconhecimento de Palavras e de Pseudopalavras. Submetido.

Vilhena, D. A., Sucena, A., Castro, S. L., & Pinheiro, Â. M. V. (2016). Reading Test-Sentence Comprehension: An Adapted Version of Lobrot's Lecture 3 Test for Brazilian Portuguese. *Dyslexia*, *22*(1), 47-63. http://dx.doi.org/10.1002/dys.1521

World Medical Association (2008). *7th Revision of the Declaration of Helsinki*: Ethical principles for medical research involving human subjects. 59th World Medical Association General Assembly, Seoul, South Corea.