

Official Organ of the European Association of Psychological Assessment

europaean journal of psychological assessment

www.hogrefe.com/journals/ejpa

Edited by
Matthias Ziegler

Abstracted/Indexed in:
Current Contents/
Social & Behavioral Sciences
Social Sciences Citation Index (SSCI)
Social Scisearch
PsycINFO
Psychological Abstracts
PSYINDEX
ERIH
Scopus

HOGREFE



Official Organ of the European Association of Psychological Assessment

europaean journal of psychological assessment

Your article has appeared in a journal published by Hogrefe Publishing. This e-offprint is provided exclusively for the personal use of the authors. It may not be posted on a personal or institutional website or to an institutional or disciplinary repository.

If you wish to post the article to your personal or institutional website or to archive it in an institutional or disciplinary repository, please use either a pre-print or a post-print of your manuscript in accordance with the publication release for your article and our “Online Rights for Journal Articles”

(www.hogrefe.com/journals).

Validity Evidence of the Test of Word Reading for Portuguese Elementary Students

Séli Chaves-Sousa,¹ Iolanda Ribeiro,¹ Fernanda Leopoldina Viana,²
Ana Paula Vale,³ Sandra Santos,¹ and Irene Cadime¹

¹School of Psychology, University of Minho, Portugal, ²Institute of Education, University of Minho, Portugal, ³Education and Psychology Department, University of Trás-os-Montes e Alto Douro, Portugal

Abstract. This article presents data concerning the validity evidence of a test of word reading (TLP – *Teste de Leitura de Palavras*) that assesses single-word reading accuracy in Portuguese students from Grades 1 to 4. The test is composed of four vertically scaled forms, one for each grade, allowing for the evaluation of students' word reading progression. In the internal structure-related evidence validity study, administration of the TLP to 905 Portuguese elementary students confirmed its one-dimensional structure. In the evidence based on the relationship with other variables' study, 280 Portuguese elementary students were assessed with each TLP test form and criterion measures (reading tests, digit span, and vocabulary measures). Significant correlations were observed between results from the TLP test forms, reading instruments, short-term memory, and vocabulary measures. These results revealed adequate evidence based on internal structure and evidence based on the relationship to other variables of the TLP for the assessment of word reading accuracy in Portuguese elementary students.

Keywords: word reading, assessment, validity, internal structure

Automatic word recognition is an essential skill in reading development (Compton et al., 2005) and a significant predictor of reading acquisition (Verhoeven & van Leeuwe, 2009). Deficits in word reading skills have been observed in students with reading disabilities (Compton & Carlisle, 1994; Perfetti, 1985). Children having deficits in accurate word reading are likely to read slowly and dysfluently, leaving fewer cognitive resources available for comprehension. Several studies report moderate to high correlations between word reading and oral reading fluency (Adlof, Catts, & Little, 2006; Meisinger, Bloom, & Hynd, 2010; Speece & Ritchey, 2005; Speece et al., 2010), as well as moderate correlations between word reading and reading comprehension (Best, Floyd, & McNamara, 2008; Ouellette, 2006; Ricketts, Nation, & Bishop, 2007). Therefore, the validation of a word reading test is justified by its relative importance for students in the early grades (Adlof et al., 2006).

Individual differences in word reading might be partially associated with differences in short-term memory (STM) and vocabulary knowledge. Swanson and Berninger (1995) found that STM is more important than working memory (WM) in predicting word reading, whereas WM is more related to reading comprehension. In addition,

moderate correlations have been found between STM and word reading (Swanson & Howell, 2001). Research has also provided substantial evidence for a low to moderate association between vocabulary and word reading (Ouellette, 2006; Ricketts et al., 2007). According to Kirby, Desrochers, Roth, and Lai (2008), increased vocabulary “allows children to recognize words they are sounding out (e.g., by mapping letter-strings onto units in the orthographic and phonological lexicons)” (p. 104) and thus facilitates word reading.

Sim-Sim and Viana (2007) conducted a systematic review of reading tests in Portugal, and concluded that the existing reading tests lacked appropriate validation procedures and standardized norms, and they omitted information about the characterization of the participants used in their development. This review had a large impact on the Portuguese scientific community and has led to increased efforts to develop new reading tests that assess: (a) reading comprehension (e.g., TCL; *Teste de Compreensão da Leitura*; Cadime et al., 2013; Cadime, Ribeiro, Viana, Santos, & Prieto, 2014); (b) oral reading fluency (e.g., TAFPL, *Teste de Avaliação da Fluência e da Precisão da Leitura “O Rei”*; Carvalho, 2010); and (c) word reading (e.g., ALEPE, *Avaliação da Leitura em Português Europeu*;

Sucena & Castro, 2012; PRP, Prova de Reconhecimento de Palavras; Viana, Ribeiro, Maia, & Santos, 2013). The word reading subtest of the ALEPE includes different lists for different grades. Therefore, the results of the different forms cannot be directly compared. In the PRP, the same items are used in the Grades 1–4 and ceiling effects after third grade were observed.

The test of word reading (TLP, Teste de Leitura de Palavras, (Viana et al., 2014)) is an original instrument designed to assess word reading accuracy, consisting of four forms (TLP-1, TLP-2, TLP-3, and TLP-4) developed for Portuguese elementary students of Grades 1–4, respectively. Each test form consists of 30 words selected from an initial pool of items that comprised 142 words which varied in terms of frequency (75 high-frequency and 67 low-frequency words), regularity (100 regular and 42 irregular words), and length (72 short and 70 long words – more than two syllables). The difference in the number of words in each category reflects the psycholinguistic characteristics of the Portuguese European language. The item selection for each test form used the one-parameter Rasch model and considered the appropriateness, in terms of difficulty, of each word to assess word reading ability of students from a particular school grade. To ensure the comparability of the results between different forms of the test, the four forms were vertically linked using common-item nonequivalent groups' design (Kolen & Brennan, 2010).

The goal of the present study was to collect evidence of validity of the test forms of the TLP based on internal structure and based on the relationship with other variables (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Kane, 2013). The first aim of this study was to provide evidence about the hypothesized one-dimensional factor structure of the TLP. This hypothesis is based on previous studies with tests that measure word reading. Athayde, Giacomoni, Zanon, and Stein (2014) found a one-dimensional structure in the study of the internal structure of the word reading subtest of the School Achievement Test. Viana and colleagues (2013) found some evidence of a one factor structure for the PRP, although the goodness-of-fit indices provided mixed results. Evidence concerning the unidimensionality of each test form of the TLP was obtained during its development using the principal component analysis of the linearized Rasch residuals (PCAR) (Viana et al., 2014).

The second aim of this study was to provide validity evidence based on relationship to other variables that research has related to word reading. The external measures assess word reading, oral reading fluency, reading comprehension, vocabulary, and STM. High correlations were expected between the scores of the test forms of the TLP and other measure that assesses word reading. Based on previous studies (Adlof et al., 2006; Best et al., 2008; Ouellette, 2006; Speece et al., 2010; Swanson & Berninger, 1995), we expected: (a) high to moderate correlations with the oral reading fluency measure; (b) moderate correlations with the reading comprehension measure; and (c) low to moderate correlations with both vocabulary and STM tasks.

Finally, teacher evaluations of students' reading skills were also used as external criteria. Research shows that teachers' judgments and reading performance are significantly and moderately correlated (Begeny, Krouse, Brown, & Mann, 2011). A similar magnitude of correlation is expected in this study.

Materials and Methods

Participants and Procedures

Data were gathered from two separate samples. Data from sample 1 were used in the internal structure-related evidence analyses and data from sample 2 were used in the analyses of evidence based on the relationship with other variables. The first sample ($n = 905$) consisted of 206 children in Grade 1 (52.8% male), 229 children in Grade 2 (52.8% male), 235 children in Grade 3 (58.8% male), and 235 in Grade 4 (46.4% male). Students attended Portuguese private (18.1%) and public schools (81.9%) from the northern and center regions of Portugal. In 2014, national data indicated that approximately 11.7% of elementary students attended private schools. Participants attended schools from different locales: urban (18.1%), suburban (35.6%), suburban (20.3%), and rural locations (21.1%). The second sample ($n = 280$) consisted of 62 first-graders (58.1% male), 66 second-graders (48.5% male), 68 third-graders (58.8% male) and 84 fourth-graders (50% male) from six urban public schools of the north of Portugal. Participants from the two samples were aged between 6 and 11 years (mean = 8.01, $SD = 1.23$), and were all native Portuguese speakers.

Data collection was authorized by the Portuguese Education Ministry, school boards, and parents or legal tutors. All measures were administered by trained psychologists between May and June of 2012 (sample 1) and May and June of 2013 (sample 2). The TLP was administered to all participants, whereas the remaining measures were administered only to the students in sample 2.

Measures

The TLP (Viana et al., 2014) assesses word reading accuracy and is composed of four vertically equated test forms. Each test form consists of 30 words. The frequency, regularity, and length of the words in each form is as follows: (a) TLP-1 – 17 high-frequency and 13 low-frequency words; 21 regular and 9 irregular words; 17 short and 13 long words; (b) TLP-2 – five high-frequency and 25 low-frequency words; 22 regular and eight irregular words; 13 short and 17 long words; (c) TLP-3 – 3 high-frequency and 27 low-frequency words; 14 regular and 16 irregular words; 11 short and 19 long words; (d) TLP-4 – 2 high-frequency and 28 low-frequency words; 17 regular and 13 irregular words; 17 short and 13 long words. The number

of low/high frequency, regular/irregular, and short/long words is not the same among the four test forms. This reflects the adopted criterion (i.e., item difficulty) in the development of the test forms, regardless of the psycholinguistic characteristics of words. The TLP is administered individually using a digital format. Children are instructed to read aloud the words that appear on the computer screen. There are no time limits for the presentation of each word or for the total time to perform the test. One point is awarded for each word read aloud accurately, and zero points are awarded for errors (omissions, substitutions, additions, or no response). The TLP has satisfactory reliability coefficients: the Person Separation Reliability (PSR), the Item Separation Reliability (ISR), and the Kuder-Richardson 20 (KR-20) coefficients are high for TLP-1 (PSR = .91; ISR = .99; KR-20 = .92), TLP-2 (PSR = .88; ISR = .99; KR-20 = .92), TLP-3 (PSR = .82; ISR = .98; KR-20 = .86), and TLP-4 (PSR = .74; ISR = .97; KR-20 = .82).

The PRP (Viana et al., 2013) is a criterion-referenced test to assess silent word reading in students from Grade 1 to 4 that can be administered individually or in groups. For students in Grades 1 and 2, the PRP has a time limit of 4 min, and for third- and fourth-graders, the time limit is 2 min. The PRP consists of 40 multiple-choice items, each of which is composed of one image, that represents a regular word (e.g., *banana* [banana]), and four written words, three of which are orthographic neighbors of the target-word (e.g., *badana*, *bacana*, *baiana*). Students have to choose the word that corresponds to the image and make a mark on it. One point is awarded for each correct response, and the total score can vary from 0 to 40 points. Cronbach's alphas for the PRP ranged between .96 and .98, test-retest reliability coefficients ranged between .76 and .88, and correlations with external criteria ranged between .36 and .62.

The TAFPL (Carvalho, 2010) assesses accuracy and oral reading fluency (ORF) in students from second to sixth grades. It consists of a narrative text with 281 words that children are instructed to read aloud. The time limit for the TAFPL is 3 min. Test-retest reliability coefficients ranged between .68 and .94 and correlations with external criteria ranged between .24 and .84.

The TCL (Cadime et al., 2013, 2014) is a norm-referenced test that assesses four components of reading comprehension (literal comprehension, inferential comprehension, reorganization, and critical comprehension) in students from second to fourth grades. The TCL includes three forms (one for each grade): TCL-2, TCL-3, and TCL-4. Each test form has 30 items. The text, which is identical across the three test forms, consists of poems, as well as narrative, informative, and instructional sequences. Items are multiple-choice with four options, only one of which is correct for each item. The reliability coefficients were .70 (PSR), .97 (ISR), and .71 (KR-20) for the TCL-2; .78 (PSR), .98 (ISR), and .79 (KR-20) for the TCL-3; and .79 (PSR), .98 (ISR), and .80 (KR-20) for the TCL-4.

The Vocabulary subtest of the WISC-III – Wechsler Intelligence Scale for Children (Wechsler, 2003) was administered and scored according to the standardized instructions from the WISC-III manual. Children have to

define a list of concepts or words. Reliability coefficients for the Portuguese Vocabulary subtest, calculated using the split-half method and corrected with the Spearman-Brown formula, ranged between .69 and .89.

The Digit Span subtest of the WISC-III – Wechsler Intelligence Scale for Children (Wechsler, 2003) was administered according to the standardized instructions for the forward digit presentation. Reliability coefficients for the Portuguese Digit Span subtest, calculated using the split-half method and corrected with the Spearman-Brown formula, ranged between .71 and .90.

The teacher evaluations of students' reading skills were assessed by asking them to rate decoding and reading comprehension skills for each of their students using a scale ranging from 1 (= *poor*) to 5 (= *excellent*).

Statistical Analyses

Confirmatory factor analysis (CFA) was conducted to investigate the hypothesized unidimensionality of the TLP, using the WLSMV estimator. According to Muthén, Du Toit, and Spisic (1997) this is a robust estimator to use with categorical data. The following fit indices were used: the Chi-Square Test (χ^2) test, in which a nonsignificant p -value indicates a good fit; the Root Mean Square Error of Approximation (RMSEA), for which values less than 0.05 indicate a good fit (Browne & Cudeck, 1993); the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI), for which values higher than 0.95 are considered to indicate a good fit (Hu & Bentler, 1999); and the Weighted Root Mean Square Residual (WRMR), for which values lower than 1.00 are considered to indicate good model fit (Yu, 2002). The analyses were performed using Mplus software version 6.1 (Muthén & Muthén, 2010).

Statistical analyses for the examination of the validity evidence based on the relationship to other variables were performed using IBM® Statistical Package for the Social Sciences®, version 20. The test scores of TLP-1, TLP-2, TLP-3, and TLP-4 were correlated with the results from the PRP, TCL, TAFPL, and Vocabulary and Digit Span subtests, as well as with the teachers' ratings. Correlations exceeding .10, .30, and .70 were considered low, moderate, and high, respectively. Age-related differences between correlations were tested according to Preacher's (2002) suggestion to convert each correlation coefficient in a z -score using Fisher's r -to- z transformation. Then, the z -scores were compared using the formula 2.8.5 from Cohen and Cohen (1983).

Results

Evidence Based on Internal Structure

Table 1 presents data concerning goodness-of-fit statistics for the TLP. Chi-square values were statistically significant, except for the TLP-4. This test is sensitive to sample size

Table 1. Global fit indicators

Form	<i>n</i>	χ^2 (405)	<i>p</i>	RMSEA	90% CI	CFI	TLI	WRMR
TLP-1	206	491.99	.002	0.032	0.021–0.042	0.982	0.981	0.959
TLP-2	229	568.53	< .001	0.042	0.034–0.050	0.966	0.964	1.067
TLP-3	235	466.38	.019	0.025	0.011–0.035	0.975	0.973	0.905
TLP-4	235	437.80	.126	0.019	0.000–0.030	0.970	0.968	0.887

Notes. RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; WRMR = Weighted Root Mean Square Residual.

Table 2. Correlations between the four forms of the TLP and the criterion measures

Form	<i>n</i>	PRP	TAFPL	TCL	Digit span	Vocabulary	Teachers' ratings	
							Decoding	Reading comprehension
TLP-1	62	.59***			.25	.23	.71***	.68***
TLP-2	66	.43***	.55***	.54***	.31*	.36*	.39**	.35**
TLP-3	68	.44**	.56***	.37**	.30*	.45**	.34**	.37**
TLP-4	84	.31**	.59***	.55***	.50***	.50***	.54***	.49***

Notes. PRP = Word Recognition Test; TAFPL = Fluency and Accuracy Assessment Test; TCL = Reading Comprehension Test. * $p < .05$. ** $p < .01$. *** $p < .001$.

and its interpretation should consider the degrees of freedom for the test (Byrne, 2012).

In the four forms of the TLP, CFI, and TLI values were higher than the minimum criterion of 0.95 and the RMSEA was lower than .05. Concerning the WRMR, values were lower than 1.00 for all TLP forms, except for TLP-2. However, the remaining fit indicators for the TLP-2 were good, such that this WRMR value alone is not sufficient to indicate bad model fit. Items' loadings varied between .418 and .895 in the TLP-1, .333 and .925 in the TLP-2, .170 and .884 in the TLP-3, and .069 and .856 in the TLP-4. All items are statistically significant indicators of the latent factor, excepting one item from the TLP-3 with the minimum item loading value of .170 and one item from the TLP-4 with the minimum item loading value of .069. Considering the overall pattern of fit statistics, all of the TLP forms showed satisfactory model fit, thereby supporting the hypothesis of a unidimensional structure.

Evidence Based on the Relationship to Other Variables

Correlations between the four forms of the TLP and the PRP, TCL, and TAFPL were positive and moderate, ranging from .31 to .59 (Table 2). Correlations between the TLP and the PRP were positive and statistically significant. Correlation differences were statistically significant only between Grade 1 and 4 ($z = 2.087$, $p = .037$), indicating a smaller correlation between the TLP and the PRP in fourth grade. In Grades 2–4 the positive correlations between the TLP and the TAFPL were also statistically significant, and the values were similar in the three grades. The correlations

with the TCL were also statistically significant and moderate in the three grades. Correlations with Vocabulary and Digit Span tasks were also statistically significant, except in Grade 1. The correlations between TLP results and teacher evaluations of decoding and reading comprehension were also positive and statistically significant. Regarding the correlations between the decoding teachers' rating and TLP, the coefficients were higher in Grade 1 than in Grade 2 ($z = 2.624$, $p = .009$) and Grade 3 ($z = 2.559$, $p = .01$). The correlations between the reading comprehension teachers' evaluation and TLP were higher in Grade 1 than in Grade 2 ($z = 2.965$, $p = .003$) and Grade 3 ($z = 2.451$, $p = .014$). No significant differences in correlations between decoding and reading comprehension teachers' evaluation and the TLP scores in Grade 1 and Grade 4 were found.

Discussion

This study presents validity evidence based on internal structure and evidence based on the relationship to other variables for the four forms of the TLP. Considering the overall pattern of fit statistics from the CFA, all the TLP forms showed satisfactory model fit, supporting the hypothesis of a unidimensional structure. This result represents an important contribution to the research on the factor structure of word reading tests, which has been limited. However, the small number of previous studies (Athayde et al., 2014; Viana et al., 2013) also suggests that replication studies are needed to confirm the data presented in this study.

The correlation coefficients between results from the TLP and the PRP were significant yet moderate. A higher association between these instruments was expected, as both tests measure word reading skills. However, the TLP involves reading words aloud and is not time limited, whereas the PRP involves silent word reading and matching an image to its correct name within a time limit. In addition, in the PRP, it is not possible to assess correct pronunciation of words. Furthermore, the TLP includes words that are both regular and irregular, whereas the PRP only includes regular ones. The reduced magnitude of correlations from first to fourth grade observed in the current study may also be explained by ceiling effects on the PRP (Viana & Ribeiro, 2010).

Correlations between the TLP and TCL were positive, of moderate magnitude, and statistically significant, consistent with previous research. Best, Floyd, and McNamara (2008) examined the influence of word reading on third-graders' reading comprehension. Correlations between three comprehension measures for narrative tests were significantly and moderately correlated with word reading ($r = .47$; $r = .46$; $r = .48$). Comparable correlations between visual word reading and reading comprehension ($r = .597$) were reported in a sample of fourth-grade students by Ouellette (2006), and in a sample of 8- and 9-year-old children (Ricketts et al., 2007). The magnitude of correlations in this study was higher for words with inconsistent spelling-sound correspondences ($r = .52$) and smaller for regular words ($r = .34$). This correlation magnitude was expected since word reading is a necessary, but not sufficient, component of reading comprehension, according to the Simple View of Reading (Hoover & Gough, 1990). In this model, reading comprehension is the product (and not the sum) of two components: word reading and oral language comprehension.

In this study, statistical significant correlations were found between ORF and word reading. Context reading fluency considerably relies on context-free word reading but it also depends on other processes emerging from the context (Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003). This would explain why correlations between the two variables found in this study and others are moderate (Speece & Ritchey, 2005; Speece et al., 2010).

Correlations between the TLP and teachers' ratings varied from moderate to high for word reading and reading comprehension, consistent with previous research. Hoge and Coladarsi (1989) reviewed studies that related teacher evaluations and students' achievement, and found that the median correlations were overall moderate in magnitude ($r = .66$) and varied from .28 to .92. Similar results were obtained in a more recent study, in which correlation coefficients between teachers' assessment and students' reading performance ranged between .43 and .58 (Begeny et al., 2011). Despite the measure and the procedures used by the teachers in their assessments may be somehow different than those used by researchers' measures (Begeny et al., 2011), teachers' assessment could be seen as valid tests of student academic performance (Gerber & Semmel, 1984).

According to models of skilled word reading (e.g., Coltheart, 2005), vocabulary knowledge is central to efficient visual word reading (Nation & Snowling, 1998). In a sample of fourth-grade students, Ouellette (2006) reported a correlation of .298 between depth of vocabulary knowledge and word reading. Moderate correlations between vocabulary and both regular ($r = .34$) and irregular words ($r = .58$) were observed by Ricketts and colleagues (2007). Correlations of the TLP with vocabulary in the present study were also positive in Grades 2–4. However, the correlation between vocabulary and the TLP-1 was not statistically significant. This result in Grade 1 was also found by Silvén, Poskiparta, Niemi, and Voeten (2007). Vocabulary only showed an indirect influence on word reading skills in first grade through the mediation effect of children's awareness of sounds. According to a developmental approach, word reading can be performed in two ways. In initial grades, children read words through a serial grapheme-phoneme conversion. Later, at about Grade 4 (Joshi & Aaron, 2000), they use the sight-word reading skill, a more automatic way to read words as a whole, retrieving them from their memory (e.g., Coltheart, 2005). This sight-word reading skill implies that children have consistent orthographic representations of the words in their mental lexicon. For the beginning reader the word representations are mainly phonological, however for the fourth-grade students word meanings are defined and added to these representations (Ouellette, 2006). At this time, the children's vocabulary knowledge becomes an important aspect of word reading since the more words they know, the more words they are able to accurately read. In the TLP-4, vocabulary knowledge becomes more relevant because it includes a larger number of irregular and low-frequency words. An accurate reading of these words implies previous knowledge.

Other than the lack of association between the TLP-1 and the digit span task, all other TLP forms were moderately and significantly correlated with STM. As with the vocabulary measure, the highest correlation was found in Grade 4, although this difference was not statistically significant from the ones obtained in the other grades. The results of this study are similar to those of other studies on STM. This variable has been associated with word reading, with moderate correlations ($r = .56$, Swanson & Howell, 2001). The study carried out by Swanson and Howell (2001) showed that older children significantly outperformed younger participants on the forward digit span task and STM predicted age differences in word reading. The authors interpreted these results as reflecting less resistance to interference, as well as difficulty in preventing the introduction of unnecessary information in younger children. These results were obtained with a sample of older students than the ones in the present study. Therefore, the relationship between the TLP and STM should be addressed in future studies.

The data from the present study demonstrated satisfactory validity evidence of the test scores of the test forms of the TLP. Identifying word reading disabled children is one of the purposes of the TLP in order to refer them to

intervention. Thus, further research to gather evidence based on consequences of testing may be useful to inform educators and researchers about validity decisions.

References

- Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing, 19*, 933–958. doi: 10.1007/s11145-006-9024-z
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Athayde, M. L., Giacomoni, C. H., Zanon, C., & Stein, L. M. (2014). Evidências de validade do subteste de leitura do teste de desempenho escolar [Evidence of validity of the reading subtest of school achievement test]. *Revista Psicologia: Teoria e Prática, 16*, 131–140. doi: 10.15348/1980-6906/psicologia.v16n2p131-140
- Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review, 40*, 23–38.
- Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29*, 137–164. doi: 10.1080/02702710801963951
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage.
- Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications and programming*. New York, NY: Routledge Academic.
- Cadime, I., Ribeiro, I., Viana, F. L., Santos, S., & Prieto, G. (2014). Calibration of a reading comprehension test for Portuguese students. *Anales de Psicologia, 30*, 1025–1034. doi: 10.6018/analesps.30.3.172611
- Cadime, I., Ribeiro, I., Viana, F. L., Santos, S., Prieto, G., & Maia, J. (2013). Validity of a reading comprehension test for Portuguese students. *Psicothema, 25*, 384–389. doi: 10.7334/psicothema2012.288
- Carvalho, A. C. (2010). *Teste de avaliação da fluência e precisão da leitura: O REI [Fluency and Accuracy Assessment Test: THE KING]*. Vila Nova de Gaia, Portugal: Edipsico.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Coltheart, M. (2005). Modeling reading: The dual-route approach. In M. J. Snowling & C. Hume (Eds.), *The science of reading: A handbook* (pp. 6–23). Oxford, UK: Blackwell Publishing.
- Compton, D. L., & Carlisle, J. F. (1994). Speed of word recognition as a distinguishing characteristic of reading disabilities. *Educational Psychology Review, 6*, 115–140. doi: 10.1007/BF02208970
- Compton, D. L., Olinghouse, N. G., Elleman, A., Vining, J., Appleton, A. C., Vail, J., & Summers, M. (2005). Putting transfer back on trial: Modeling individual differences in the transfer of decoding-skill gains to other aspects of reading acquisition. *Journal of Educational Psychology, 97*, 55–69. doi: 10.1037/0022-0663.97.1.55
- Gerber, M. M., & Semmel, M. I. (1984). Teacher as imperfect test: Reconceptualizing the referral process. *Educational Psychologist, 19*, 137–148. doi: 10.1080/00461528409529290
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*, 297–313. doi: 10.3102/00346543059003297
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127–160. doi: 10.1007/BF00401799
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55.
- Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*, 719–729. doi: 10.1037/0022-0663.95.4.719
- Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: simple view of reading made a little more complex. *Reading Psychology, 21*, 85–97. doi: 10.1080/02702710050084428
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1–73. doi: 10.1111/jedm.12000
- Kirby, J. R., Desrochers, A., Roth, L., & Lai, S. S. V. (2008). Longitudinal predictors of word reading development. *Canadian Psychology/Psychologie Canadienne, 49*, 103–110. doi: 10.1037/0708-5591.49.2.103
- Kolen, M. J., & Brennan, R. L. (2010). *Test equating, scaling and linking* (2nd ed.). New York, NY: Springer.
- Meisinger, E. B., Bloom, J. S., & Hynd, G. W. (2010). Reading fluency: Implications for the assessment of children with reading disabilities. *Annals of Dyslexia, 60*, 1–17. doi: 10.1007/s11881-009-0031-z
- Muthén, B. O., Du Toit, S. H., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. (Unpublished technical report). Retrieved from http://www.gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf
- Muthén, B. O., & Muthén, L. (2010). Mplus Version 6.1 [Software]. Los Angeles, CA: Muthén & Muthén.
- Nation, K., & Snowling, M. J. (1998). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development, 69*, 996–1011. doi: 10.1111/j.1467-8624.1998.tb06157.x
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*, 554–566. doi: 10.1037/0022-0663.98.3.554
- Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford University Press.
- Preacher, K. J. (2002). Calculation of the test of the difference between two independent correlation coefficients [Computer software]. Available in <http://quantpsy.org>
- Ricketts, J., Nation, K., & Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading, 11*, 235–257. doi: 10.1080/10888430701344306
- Silvén, M., Poskiparta, E., Niemi, P., & Voeten, M. (2007). Precursors of reading skill from infancy to first grade in Finnish: Continuity and change in a highly inflected language. *Journal of Educational Psychology, 99*, 516–531. doi: 10.1037/0022-0663.99.3.516
- Sim-Sim, I., & Viana, F. L. (2007). *Para a avaliação do desempenho de leitura [To evaluate reading achievement]*. Lisboa, Portugal: Ministério da Educação – Gabinete de Estatística e Planeamento da Educação.
- Speece, D. L., & Ritchey, K. D. (2005). A longitudinal study of the development of oral reading fluency in young children at

- risk for reading failure. *Journal of Learning Disabilities* (pp. 387–399). doi: 10.1177/00222194050380050201
- Speece, D. L., Ritchey, K. D., Silverman, R., Schatschneider, C., Walker, C. Y., & Andrusik, K. N. (2010). Identifying children in middle childhood who are at risk for reading problems. *School Psychology Review*, 39, 258–276.
- Sucena, A., & Castro, S. L. (2012). *ALEPE – Avaliação da leitura em Português Europeu* [ALEPE – Reading Assessment in European Portuguese]. Lisboa, Portugal: CEGOC.
- Swanson, H. L., & Berninger, V. (1995). The role of working memory in skilled and less skilled readers' comprehension. *Intelligence*, 21, 83–108. doi: 10.1016/0160-2896(95)90040-3
- Swanson, H. L., & Howell, M. (2001). Working memory, short-term memory, and speech rate as predictors of children's reading performance at different ages. *Journal of Educational Psychology*, 93, 720–734. doi: 10.1037/0022-0663.93.4.720
- Verhoeven, L., & van Leeuwe, J. (2009). Modeling the growth of word-decoding skills: Evidence from Dutch. *Scientific Studies of Reading*, 13, 205–223. doi: 10.1080/10888430902851356
- Viana, F. L., & Ribeiro, I. (2010). *PRP- Prova de Reconhecimento de palavras. Manual Técnico* [Word Recognition Test. Technical Manual]. Lisboa, Portugal: CEGOC.
- Viana, F. L., Ribeiro, I., Maia, I., & Santos, S. (2013). Propriedades psicométricas da Prova de Reconhecimento de Palavras [Psychometric properties of the Word Recognition Test]. *Psicologia: Reflexão e Crítica*, 26, 231–240. doi: 10.1590/S0102-79722013000200003
- Viana, F. L., Ribeiro, I., Vale, A. P., Chaves-Sousa, S., Santos, S., & Cadime, I. (2014). *TLP – Teste de Leitura de Palavras. Manual Técnico* [Test of Word Reading, Technical Manual]. Lisboa: CEGOC-TEA.
- Wechsler, D. (2003). *Escala de Inteligência de Wechsler para Crianças – Terceira edição (WISC-III)* [Wechsler Intelligence Scale for Children – Third Edition]. Lisboa, Portugal: CEGOC.
- Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Los Angeles, CA: University of California.

Date of acceptance: June 25, 2015

Published online: December 22, 2015

Séli Chaves-Sousa

Centro de Investigação em Psicologia
Universidade do Minho
Campus de Gualtar
4710-057 Braga
Portugal
Tel. +35 19 3628-8826
E-mail seli.chaves.sousa@gmail.com
