Cognition and Neurosciences

Assessing reading comprehension with narrative and expository texts: Dimensionality and relationship with fluency, vocabulary and memory

SANDRA SANTOS, IRENE CADIME, FERNANDA L. VIANA, SÉLI CHAVES-SOUSA, ELENA GAYO, JOSÉ MAIA and IOLANDA RIBEIRO

1School of Psychology, University of Minho, Braga, Portugal
2Institute of Education, University of Minho, Braga, Portugal
3University of Vigo, Pontevedra, Portugal
4Faculty of Sport, University of Porto, Porto, Portugal


Reading comprehension assessment should rely on valid instruments that enable adequate conclusions to be taken regarding students’ reading comprehension performance. In this article, two studies were conducted to collect validity evidence for the vertically scaled forms of two Tests of Reading Comprehension for Portuguese elementary school students in the second to fourth grades, one with narrative texts (TRC-n) and another with expository ones (TRC-e). Two samples of 950 and 990 students participated in Study 1, the study of the dimensionality of the TRC-n and TRC-e forms, respectively. Confirmatory factorial analyses provided evidence of an acceptable fit for the one-factor solution for all test forms. Study 2 included 218 students to collect criterion-related validity. The scores obtained in each of the test forms were significantly correlated with the ones obtained in other reading comprehension measures and with the results obtained in oral reading fluency, vocabulary and working memory tests. Evidence suggests that the test forms are valid measures of reading comprehension.

Key words: Reading comprehension, reading assessment, validity, confirmatory factorial analysis, narrative texts, expository texts.

INTRODUCTION

The concept of reading comprehension is understood as the ability to extract and construct meaning from written language, which implies a close interaction between the reader and the text’s characteristics (RAND Reading Study Group, 2002) and involves the acquisition and the development of reading related skills such as reading fluency and vocabulary. Given that reading comprehension is one of the areas in which students struggle and its impact on the several academic subjects may be very significant, reading comprehension measures should allow an accurate assessment providing the possibility of comparing student’s performance with that of a normative group (RAND Reading Study Group, 2002) and the study of the students’ intra-individual changes and inter-individual differences in performances across grade levels (Cain, 2010).

In this article, we present validity evidence for two tests of reading comprehension for Portuguese students in elementary school: the TRC-n – Test of Reading Comprehension of Narrative Texts and the TRC-e – Test of Reading Comprehension of Expository Texts. Each test is composed of three vertically scaled forms (TRC-n-2, TRC-n-3 and TRC-n-4; TRC-e-2, TRC-e-3 and TRC-e-4), generated through Rasch model analyses (Santos, Cadime, Viana et al., 2016). These test forms were constructed to assess the same construct, and were designed to be as similar as possible in content but were allowed to vary in terms of difficulty, since these tests measure a construct (reading comprehension) that changes with learning and is expected to increase with academic grade levels (Kolen & Brennan, 2010). Therefore, using a vertical scaling procedure, the scores of each test form were adjusted to a common scale, enabling the comparison of performance in reading comprehension measured with the different test forms (Santos et al., 2016). The first study in this article is focused on validity evidence based on the internal structure of each test form and the second one is dedicated to the validity evidence based on relations with other variables, namely reading fluency, vocabulary and working memory.

Reading comprehension: The construct and its assessment

Due to the importance of reading across the curriculum and grade levels, in the construction of reading comprehension tests a set of features to reach the goal of a more appropriate assessment that enables the identification of students with reading comprehension difficulties and the guidance of teaching practices according to the students’ needs have been considered. Text related factors have been considered, for example, by including texts of different genres in reading comprehension tests. This has been the option, for example, in the reading comprehension subtest of the Gates–MacGinitie Reading Test (GMRT) (MacGinitie, MacGinitie, Maria, Dreyer & Hughes, 2002), in a Portuguese reading comprehension test – TCL-Teste de Compreensão da Leitura (Cadime, Ribeiro, Viana, Santos & Prieto, 2014; Cadime, Viana & Ribeiro, 2013), and in the Spanish reading comprehension tests ACL-Pruebas de Evaluación de la Comprensión Lectora (Catalá, Catalá, Molina & Monclús, 2001) and ECOMPLEC- Evaluación de la Comprensión Lectora (León, Escudero & Olmos, 2012). This option is based on research findings that support that text
genre plays an important role in comprehension as performance on reading comprehension can be different when narrative or expository texts are used (Eason, Goldberg, Young, Geist & Cutting, 2012; Mullis, Martin, Foy & Drucker, 2012; RAND Reading Study Group, 2002). Individual differences in performance have been explained by the inherent differences between the two genres of text with respect to their purpose (Duran, McCarthy, Graesser & McNamara, 2007), structure (Best, Floyd, McNamara & Danielle, 2008), vocabulary (Cain & Oakhill, 2006), as well as the necessary previous knowledge to generate inferences and elaborate a coherent mental representation of the texts (Wolfe & Woody, 2010).

In addition, tests of reading comprehension have been built to cover different levels of reading comprehension (Basaraba, Yovanoff, Alonzo & Tindal, 2013; Eason et al., 2012), which refer to the demands of the comprehension tasks that are proposed to the reader concerning the complexity of the information processing and the use of different sources of information to complete a specific comprehension task. Comprehension levels support not only the definition of pedagogical strategies, but also the categorization of the questions in reading comprehension assessment (Alonzo, Basaraba, Tindal & Currieaue, 2009; Dewitz, Dewitz, Mark & Haskell, 2003). Those levels can be synthesized as follows: (1) literal comprehension (LC) – to comprehend explicit information in the text; (2) inferential comprehension (IC) – to draw conclusions about information that is implicitly stated in the text; (3) reorganization (R) – to synthesize or schematize the text content and; (4) critical comprehension (CC) – to make judgments or give opinions about the text (Herber, 1978; Swaby, 1989). The different levels only specify the varying demands of the reading comprehension tasks and they are all part of the reading comprehension construct (Basaraba et al., 2013; Ozuru, Rowe, O’Reilly & McNamara, 2008).

The hypothesis of reading comprehension as a unitary construct is supported in several tests, such as the reading comprehension subtest of the Gates-MacGinitie Reading Test (GMRT) (Cook, Eignor, Steinberg, Sawaki & Clíne, 2009), where reading comprehension is assessed with a variety of text materials, through silent text reading and multiple-choice questions that demand the comprehension of explicit and implicit information. Confirmatory factorial analyses support for reading comprehension as a single dimension has been evidenced with a sample of students without disabilities and a sample of students with reading-based learning disabilities. The unidimensionality of the reading comprehension construct is also supported in the reading comprehension subtest of the Nelson-Denny Reading Ability Test (Brown, Fishco & Hanna, 1993), that also relies on multiple-choice questions to test what the authors label direct comprehension skills and the ability to make inferences based on texts’ reading, through the silent reading of a variety of text materials. Evidence for a single dimension construct is also provided for the reading comprehension subtest of the Scholastic Aptitude Test-SAT, wish shares similarities with the two subtests previously mentioned regarding its structure. Confirmatory factorial analyses support that a one-factor model fits the data well in the SAT (Dorans & Lawrence, 1999). Confirmatory factorial analyses also sustain reading comprehension as a unitary construct in the ECOMPLEC (León et al., 2012), a test that assesses reading comprehension through silent texts’ reading and the use of multiple-choice questions that cover the assessment of different comprehension levels and that enables the delivery of results for narrative and expository texts.

Finally, the hypothesis of the unidimensionality is supported via confirmatory factorial analyses in the TCL (Cadime et al., 2013), a test with vertically scaled test forms, constructed for second to fourth grade students in elementary school, that aims the assessment of the four levels of comprehension, with multiple-choice items. Despite the similarities with the TRC-n and the TRC-e, the TCL offers four alternative response options and the TRC-n and the TRC-e provide three answer choices, which is time saving in the tests’ administration. In addition, and more important, the TCL does not allow the comparison of the performance in the comprehension of narrative and expository texts. This limitation is overcome with the TRC-n and the TRC-e. The characterization of performance according to the text genre, narrative or expository, distinguishes these tests from others that integrate simultaneously both text genres such as TCL. This aspect can have a major impact on the information available to educators and psychologists about students’ abilities to comprehend a different range of texts with which students are confronted in schools (Morsy, Kieffer & Snow, 2010). Besides, the construction of vertically scaled test forms allows growth in reading comprehension to be measured from second to fourth grade (Santos et al., 2016).

**Reading comprehension: Relation to other variables**

Individual differences in reading comprehension can be associated with differences in oral reading fluency, that is, the ability to read a text with speed and accuracy. A student with higher reading fluency skills end up reading more and, as a result, reading skills are improved (Fuchs, Fuchs, Hosp & Jenkins, 2001). For this reason, moderate to high correlations have been reported between oral reading fluency and reading comprehension. Evidence of a high correlation between reading comprehension and oral reading fluency is provided by Fuchs and colleagues (2001). In their study with middle school and junior high school students with reading disabilities they found a correlation of 0.91 between oral reading fluency and reading comprehension as measured by the reading comprehension subtest of the Stanford Achievement Test. Yovanoff, Duesbery, Alonzo and Tindal (2005) found moderate correlations between oral reading fluency and reading comprehension that ranged from 0.42 to 0.65 in five cross-sectional samples of students from fourth to eighth grades. These results are in accordance with the ones obtained by Padelidou and Antoniou (2014). These authors found correlations between oral reading fluency and reading comprehension that ranged from 0.36 to 0.47 in a sample of Greek students from first to fourth grade. Moderate correlations were also obtained by Ribeiro, Cadime, Freitas and Viana (2016) in a study with Portuguese students from second and fourth grades. In this study, the correlation in the second grade reached 0.67 and in the fourth grade it was of 0.26.

A close relationship between vocabulary and reading comprehension has also been systematically observed (Cain, 2010; Cain, Oakhill & Bryant, 2004b; Carver, 2003; Joshi, 2005;
Stahl & Nagy, 2006). Students with poor vocabulary struggling with reading comprehension usually avoid reading and, therefore, learn less new words and have more difficulties comprehending what they read. In contrast, students with richer vocabulary read more, so they enhance their reading comprehension abilities (Joshi, 2005). Prior research has demonstrated moderate to high correlations between vocabulary and reading comprehension. A study by Ouellette (2006) with fourth graders showed correlations between reading comprehension and vocabulary breadth and depth that varied among 0.36 and 0.50. A further study by Ouellette and Beers (2009) provided evidence for significant correlations among reading comprehension and vocabulary measures of breadth and depth in first and six graders. The correlation coefficients assumed values of 0.49 and 0.42 in the first grade and values of 0.65 and 0.51 in the sixth grade. Evidence for significant correlations between vocabulary and reading comprehension, weather narrative or expository are used, has also been provided with fifth graders (Yildirim, Yildiz & Ates, 2011). A more recent study conducted by Ribeiro and colleagues (2016) with Portuguese students reported correlation coefficients among vocabulary and reading comprehension results of 0.47 in the second grade, and 0.35 in the fourth grade.

Reading performance can also be influenced by memory skills (Cain, Bryant & Oakhill, 2004a; Goff, Pratt & Ong, 2005; Oakhill, Cain & Bryant, 2003). Literature reports low to moderate correlations between reading comprehension and memory skills, being higher when the memory assessment tasks demand the storage and processing of information (numbers, words, phrases) (Swanson & Berninger, 1995). Statistical significant correlations between working memory and reading comprehension were found in the study of Seigneuric, Ehrlich, Oakhill and Yuill (2000) where five working memory tasks were applied to fourth grade students, two verbal, two numerical and one spatial. The correlations between working memory tasks, except the spatial task, and reading comprehension were significant, varying from 0.41 to 0.56. The study of Cain et al. (2004a) with children aged between eight and 11 years old also reported statistical significant correlations between reading comprehension and memory assessed through a digit span task, when children were eight \( (r = 0.37) \) and nine \( (r = 0.34) \) years old. The correlation coefficient was no longer statistically significant by the age of 11. In the study by Ribeiro and colleagues (2016), correlation results among working memory and reading comprehension were moderate both in the second \( (r = 0.38) \) and fourth \( (r = 0.42) \) grades.

The present study
This article aims to gather empirical validity evidence for the TRC-n and the TRC-e 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). To achieve this purpose, two studies were conducted. Study 1 aims to gather validity evidence based on the internal structure of each test form via confirmatory factorial analysis (CFA). Based on evidence of the unidimensionality of the construct, a one-factor model was postulated for the TRC-n and TRC-e.

Study 2 intends to collect validity evidence based on relations with other variables, such as reading fluency, vocabulary and working memory. Taking into account the state of the art, a close relationship between reading comprehension and these variables is expected.

STUDY 1: METHOD

Participants
In study 1, two different samples were used: one for the study of the dimensionality of each test (TRC-n or TRC-e). A total of 950 students from second \( (n = 305) \), third \( (n = 329) \) and fourth grade \( (n = 316) \) participated in the study of the TRC-n dimensionality. In each grade, the majority of the participants were male, representing 51.5% \( (n = 157) \) of the second graders, 52.6% \( (n = 173) \) of the third graders and 54.7% \( (n = 173) \) of the fourth graders. The sample used for the study of the dimensionality of the TRC-e included 990 students from the second \( (n = 329) \), third \( (n = 310) \) and fourth grade \( (n = 351) \). Male participants represented 47.7% \( (n = 157) \) of the second grade sample, 52.6% \( (n = 163) \) of the third grade sample and 51.3% \( (n = 180) \) of the fourth grade sample. All students attended public schools in the north of Portugal and had European Portuguese as their native language and none had permanent special education needs.

Instruments
The Test of Reading Comprehension of Narrative Texts (TRC-n) and the Test of Reading Comprehension of Expository Texts (TRC-e) assesses, respectively, reading comprehension of narrative and expository texts. Each test is composed of three vertically scaled forms, each one aimed at assessing students from the second, third and fourth grades of elementary school: TRC-n-2, TRC-n-3, TRC-n-4 and TRC-e-2, TRC-e-3, TRC-e-4. The number identifies the grade level. Every form of the TRC-n and the TRC-e is composed of a booklet with the texts and a worksheet with the items. The texts are original, unpublished and written, upon request, by Portuguese authors of children’s literature and scientific publications for inclusion in these tests. Items are multiple-choice with three options (one correct) and each item assesses literal comprehension (LC), inferential comprehension (IC), reorganization (R) or critical comprehension (CC). Each form of the TRC-n consists of 27 items. The number of items in each TRC-n form assessing each comprehension level is as follows: TRC-n-2: LC = 8, IC = 14, R = 3, and CC = 2; TRC-n-3: LC = 6, IC = 15, R = 2, and CC = 4; TRC-n-4: LC = 6, IC = 15, R = 2, and CC = 4. The TRC-e forms are composed of 33 items each. Each TRC-e for comprises a number of items assessing each comprehension level: TRC-e-2: LC = 15, IC = 14, R = 2, and CC = 2; TRC-e-3: LC = 10, IC = 17, R = 5, and CC = 1; TRC-e-4: LC = 8, IC = 15, R = 7, and CC = 3. All test forms demonstrate high reliability coefficients (person separation reliability – PSR, item separation reliability – ISR and the Kuder-Richardson formula 20 – KR20) ranging from 0.70 and 0.96 on the TRC-n forms and from 0.72 to 0.95 on the TRC-e forms.
Procedure

Legal authorizations for the administration of the tests were obtained from the Portuguese Ministry of Education and the school boards, and informed consent for student participation was acquired from students’ parents or legal tutors. Tests were administered collectively during regularly scheduled class time by trained psychologists. Students were instructed to read the texts and the items silently and then to mark the correct answer. No time limit was set, and no additional clarification or help was given during the administration of the tests.

Data analyses

The hypothesis of a one-factor model was tested through CFA to obtain fit statistics for each form of the TRC-n and the TRC-e results. The software used was Mplus 6.1 (Muthén & Muthén, 2010). Weighted least squares with mean and variance adjusted (WLSMV) was used as a parameter estimation method due to its robustness in handling categorical data (Brown, 2006).

A chi-square test of model fit ($\chi^2$) and four descriptive fit indices were also gathered: (1) the root-mean-square error of approximation (RMSEA); (2) the comparative fit index (CFI); (3) the Tucker-Lewis index (TLI); and (4) the weighted root mean square residual (WRMR). The chi-square test of model fit index tests the discrepancy between the unrestricted sample covariance matrix and the restricted covariance matrix (Byrne, 2011). The value of the chi-square statistic indicates a good fit when the probability associated is non-significant, that is, superior to 0.05. However, this test is sensitive to sample size so it will almost always be significant with large samples (Harrington, 2009). The RMSEA estimates the extent to which the model fits reasonably well in reproducing the population covariances (Thompson, 2004). A value below 0.05 indicates a good fit (Brown & Cudeck, 1993). The CFI and the TLI are incremental relative fit indices that measure the improvement in the model fit over a baseline model in which it is assumed that variables are uncorrelated (Kline, 2011). Values higher than 0.90 are indicators of acceptable model fit (Bentler, 1990). The WRMR, which measures the average differences in variances and covariances between the sample and the estimated population, is an appropriate index for binary data when the samples exceed 250 (Brown, 2006). A cut-off value close to 1.0 is suitable for a good fit (Yu, 2002).

RESULTS

The results of the one-factor model as tested for each form of the TRC-n and the TRC-e, thus supporting reading comprehension as a unitary construct.

### Table 1. Global fit indices

<table>
<thead>
<tr>
<th>Test Form</th>
<th>$\chi^2$</th>
<th>df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>TLI</th>
<th>WRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRC-n-2</td>
<td>351.81</td>
<td>324</td>
<td>0.017</td>
<td>0.98</td>
<td>0.97</td>
<td>0.878</td>
</tr>
<tr>
<td>TRC-n-3</td>
<td>397.92*</td>
<td>324</td>
<td>0.026</td>
<td>0.90</td>
<td>0.89</td>
<td>0.983</td>
</tr>
<tr>
<td>TRC-n-4</td>
<td>370.18*</td>
<td>324</td>
<td>0.021</td>
<td>0.95</td>
<td>0.95</td>
<td>0.920</td>
</tr>
<tr>
<td>TRC-e-2</td>
<td>552.90*</td>
<td>495</td>
<td>0.019</td>
<td>0.96</td>
<td>0.95</td>
<td>0.927</td>
</tr>
<tr>
<td>TRC-e-3</td>
<td>559.19*</td>
<td>495</td>
<td>0.020</td>
<td>0.95</td>
<td>0.94</td>
<td>0.946</td>
</tr>
<tr>
<td>TRC-e-4</td>
<td>753.44*</td>
<td>495</td>
<td>0.039</td>
<td>0.89</td>
<td>0.88</td>
<td>1.144</td>
</tr>
</tbody>
</table>

Notes: *$p < 0.05$; df = degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; WRMR = weighted root mean square residual.

other test form models exhibited WRMR values near 1.0. Data regarding the goodness-of-fit indices suggest a good fit for the one-factor solution on every test form, thus supporting reading comprehension as a unitary construct.

STUDY 2: METHOD

Participants

A group of 218 participants was used in the study of the criterion-related evidence of validity of the TRC-n and the TRC-e. Sixty-six were second graders, 68 were third graders and 84 were fourth graders. The sample distribution by gender was similar in every grade: male students represented 48.5% ($n = 32$) of the second grade sample, 58.8% ($n = 40$) of the third grade sample and 50% ($n = 42$) of the fourth grade sample. All students attended public schools and had European Portuguese as their native language and none had permanent special education needs.

Instruments

**Test of Reading Comprehension of Narrative Texts (TRC-n) and the Test of Reading Comprehension of Expository Texts (TRC-e).** See description in study 1.

**TCL Reading comprehension test** (Cadime, Viana & Ribeiro, 2014; Cadime et al., 2013). This test is composed of three vertically scaled forms to assess reading comprehension of second, third and fourth graders enrolled in elementary school. The text is common to every form, and it incorporates poems and narrative, expository and instructional sequences. Items are multiple-choice with four alternatives (one correct) and aim to assess LC, IC, R and CC. Each form consists of 30 items. The test administration is performed collectively and with no time limit. The three forms present high reliability coefficients (PSR, ISR and KR20) ranging from 0.70 to 0.98. In terms of relations to other variables, the correlations between the TCL results and the ones obtained in other reading comprehension tests were of 0.23, in the second grade, 0.66 in the third grade and 0.73, in the fourth grade. The correlations of the TCL forms’ results with teachers’ evaluation were moderate, ranging from 0.55 to 0.68.

**Reading Fluency Test** (Teste “O Rei”, Carvalho, 2010). This test measures reading fluency through a text that is read aloud by the student. Its administration is individual and its time limit is three minutes. At the end of this period, the number of correct
words is counted and an index of fluency is computed by counting the number of words correctly read per minute. This test has good reliability (0.94) and validity evidence based on relations with other variables. Specifically, the correlations between fluency and a reading comprehension measure were of 0.60 in the second grade, 0.62 in the third grade and 0.64 in the fourth grade. Correlation coefficients with teachers’ ratings of students’ performance regarding oral reading fluency were high in the second \((r = 0.72)\), third \((r = 0.69)\) and fourth \((r = 0.70)\) grades.

**Vocabulary** (Wechsler, 2003). This is a subtest of the Wechsler Intelligence Scale for Children (WISC-III). This subtest is individually administered and assesses verbal concept formation by asking students to provide a definition of a given word. This subtest has 30 items that can be scored with zero, one or two points, depending on the quality of the response, which means that the maximum raw score is 60 points. The administration is interrupted after four consecutive failures. The Portuguese version of the WISC-III presents high fidelity and good indicators of validity (Simões & Albuquerque, 2002).

**Digit Span** (Wechsler, 2003). This subtest is also part of the WISC-III and aims to assess working memory. A series of numbers with two to nine digits are presented to the students who are then asked to repeat them in the direct (forward digit span) or the reverse order (backward digit span). Each series is composed of two rehearsals. The test is interrupted after failure in both rehearsals of the same series. The administration of this subtest is individual.

**Teachers’ ratings of students’ reading comprehension and oral reading fluency skills.** Teachers rated their students on a scale ranging from 1 (very poor) to 5 (excellent) based on their performances on tasks that they proposed to their students during the school year. Despite some limitations, there is evidence for the use of teachers’ ratings as validity measures since teachers have been proved to be accurate evaluators of the students’ performance in reading comprehension and oral reading fluency measures (Feinberg & Shapiro, 2003).

**Procedure**

For the administration of the tests, legal authorizations were obtained from the Portuguese Ministry of Education and the school boards. Students’ parents or legal tutors were asked to consent for their child’s participation in the study. The TRC-n, the TRC-e and the reading comprehension test were administered collectively during regularly scheduled class time. The application of the tests of reading fluency, vocabulary and digit span was performed individually. The administration of the tests was ensured by trained psychologists.

### Data analyses

Pearson correlation coefficients were computed to analyse the relationship between the TRC-n and the TRC-e forms’ results. The common variance shared by the two tests in each grade was computed by squaring r (McGuinness, 2005).

Pearson correlation coefficients were also calculated to analyse the relationship between the TRC-n and the TRC-e forms’ results and the other measures of reading comprehension (reading comprehension test, reading fluency, vocabulary, memory, and the teachers’ evaluation of students’ oral reading fluency and reading comprehension skills. The IBM SPSS Statistics 20 was used.

### RESULTS

The analysis of the correlation coefficients between the TRC-n and the TRC-e forms’ results indicate statistically significant correlations in the second \((r = 0.58, p < 0.001)\), third \((r = 0.58, p < 0.001)\), and fourth grades \((r = 0.76, p < 0.001)\). By squaring \(r\), the results show that the TRC-n and the TRC-e share 34% of common variance in the second and third grades and 58% of common variance in the fourth grade.

Table 2 shows the descriptive statistics for the results in the TRC-n and TRC-e and the correlations between the TRC-n and the TRC-e forms and the results from the tests used as external criteria.

Positive and statistically significant correlations were obtained between all the forms of the TRC-n and TRC-e and the measures used as external criteria. Moderate to high correlations, ranging from 0.56 to 0.72, were found between the TRC-n and TRC-e forms and the reading comprehension test (Cadime et al., 2013). The correlations between the TRC-n and TRC-e forms and reading fluency were moderate, ranging from 0.39 to 0.56, and the correlations with vocabulary were low to moderate, ranging from 0.25 to 0.49. Low to moderate correlations were observed between the TRC-n and the TRC-e and memory, ranging from 0.26 to 0.51. Low to moderate correlations were found between the TRC-n and TRC-e forms and the teachers’ evaluations.

<table>
<thead>
<tr>
<th>Test form</th>
<th>(N)</th>
<th>M</th>
<th>SD</th>
<th>TCL</th>
<th>Reading Fluency Test</th>
<th>Vocabulary</th>
<th>Digit Span</th>
<th>ORF by teachers(^{a})</th>
<th>RC by teachers(^{b})</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRC-n-2</td>
<td>66</td>
<td>16.73</td>
<td>4.71</td>
<td>0.57***</td>
<td>0.48***</td>
<td>0.38**</td>
<td>0.34**</td>
<td>0.51***</td>
<td>0.55***</td>
</tr>
<tr>
<td>TRC-n-3</td>
<td>66</td>
<td>15.64</td>
<td>3.99</td>
<td>0.64***</td>
<td>0.39**</td>
<td>0.31*</td>
<td>0.26*</td>
<td>0.36**</td>
<td>0.41**</td>
</tr>
<tr>
<td>TRC-n-4</td>
<td>85</td>
<td>17.05</td>
<td>4.92</td>
<td>0.72***</td>
<td>0.48***</td>
<td>0.28*</td>
<td>0.51***</td>
<td>0.46***</td>
<td>0.36**</td>
</tr>
<tr>
<td>TRC-e-2</td>
<td>66</td>
<td>20.82</td>
<td>5.76</td>
<td>0.69***</td>
<td>0.56***</td>
<td>0.49***</td>
<td>0.43***</td>
<td>0.62***</td>
<td>0.68***</td>
</tr>
<tr>
<td>TRC-e-3</td>
<td>66</td>
<td>18.55</td>
<td>5.04</td>
<td>0.56***</td>
<td>0.41**</td>
<td>0.39**</td>
<td>0.31**</td>
<td>0.40**</td>
<td>0.60***</td>
</tr>
<tr>
<td>TRC-e-4</td>
<td>83</td>
<td>20.82</td>
<td>5.62</td>
<td>0.72***</td>
<td>0.52***</td>
<td>0.28*</td>
<td>0.41***</td>
<td>0.34**</td>
<td>0.28*</td>
</tr>
</tbody>
</table>

*Notes: \(*p < 0.05; **p < 0.01; ***p < 0.001\). \(^{a}\)Oral Reading Fluency assessed by teachers. \(^{b}\)Reading Comprehension assessed by teachers."
regarding students’ reading skills. The correlations with oral reading fluency skills ranged from 0.34 to 0.62, and the correlations with reading comprehension ranged between 0.28 and 0.68.

DISCUSSION AND CONCLUSIONS

The goals of both of the studies presented in this article were to gather evidence based on internal structure and criterion-related evidence of validity for the vertically scaled forms of the TRC-n and TRC-e, developed to assess Portuguese students enrolled in elementary school grades second to fourth. Validity studies that support the conclusions drawn from the administration of the tests are critical as they provide a basis for understanding the construct being measured (Morsy et al., 2010).

Regarding study 1, a one-factor model was tested using CFA. The results show that the chi-square statistics criteria were only met in the TRC-n-2; however, because this indicator is sensitive to sample size, it should not be considered the main indicator of model fit (Harrington, 2009). The RMSEA indices for all test forms are below the reference value, while the CFI and TLI indices are within or above the recommended values. All test forms’ models exhibited WRMR values around 1.0. Therefore, all goodness-of-fit indices provide evidence of a good fit for the one-factor solution on every test form.

Concerning study 2, the analysis of the correlations between the TRC-n and TRC-e test forms’ results indicates that these share a fair amount of common variance in the three grade levels. This result demonstrates that the scores are not excessively correlated, suggesting that each instrument can contribute differently to characterize students’ reading performance.

The correlations of the TRC-n and TRC-e with the TCL, the other reading comprehension test (Cadime et al., 2013), are moderate, which is as expected considering that these tests measure the same construct and their theoretical and methodological assumptions are similar. The magnitude of the correlations between these tests is similar to the one obtained in the study of Cadime and colleagues (2013) where the authors correlated their test with another reading comprehension test (Cadime et al., 2013) with respect to reading comprehension and vocabulary.

The results obtained in the second and third grades regarding the correlation between reading comprehension and vocabulary in the fourth grade are low, both when narrative (r = 0.28) or expository texts (r = 0.25) are used. Higher correlations were expected in accordance with the assumption of a close relationship between vocabulary and reading comprehension (Carver, 2003; Joshi, 2005; Stahl & Nagy, 2006; Stanovich, 2000). These results in the fourth grade are not consistent with the findings of studies such as the ones developed by Ouellette (2006) and Ouellette and Beers (2009), where the results suggest moderate correlations between the two variables. The results obtained in the second and third grades regarding the correlation between vocabulary and reading comprehension are similar to the ones provided by the study of Ribeiro and colleagues (Ribeiro et al., 2016).

Our findings suggest low correlations between the TRC-n and the TRC-e forms and memory as assessed by the digit span test. The values obtained match the ones obtained in other studies correlating reading comprehension with a working memory task based on numbers (Daneman & Merikle, 1996; Goff et al., 2005) and are consistent with literature that relates reading comprehension with numerical working memory (Oakhill et al., 2003; Seigneuric et al., 2000).

Low to moderate correlations were found between the TRC-n and TRC-e forms and the teachers’ evaluation with respect to oral reading fluency and reading comprehension. The magnitude of the correlation coefficients is lower than the one obtained in the study of Feinberg and Shapiro (2003) regarding oral reading fluency and lower than the results gathered in the study of Cadime and colleagues (2013) with respect to reading comprehension evaluation.

A limitation of this research concerns the construct-related validity study: although literature is consistent on supporting reading comprehension as a unidimensional construct, it would be a great contribution for research if a competing model with four factors, each corresponding to a comprehension level (literal, inferential, reorganization and critical) was tested. However, the number of items in the critical and reorganization levels is not sufficient to perform such analyses. In future investigations, tests with a similar number of items measuring each comprehension level should be analysed regarding the hypothesis of a four-factor model and the results obtained compared with the fit for a one-factor model.

In conclusion, the results of these studies suggest that the TRC-n and the TRC-e forms are valid measures of the reading comprehension of Portuguese students in the second to fourth grades of elementary school as related to, respectively, narrative and expository texts. The validity studies that were conducted support a unidimensional structure of each test form and a reliable interpretation of a single test score as a representation of the reading comprehension construct. The TRC-n and TRC-e forms allow a valid assessment of reading comprehension competencies of each student and a reliable comparison of the results to the expectations for each grade level once the raw scores are converted to percentiles and standardized scores.
The TRC-n and TRC-e emerge as two original instruments that fill a gap in Portugal, where there are very few tests to assess comprehension. Each test has three vertically scaled test forms that allow the comparison of inter and intra-individual results throughout the elementary educational course once the test forms are placed in the same metrics. This comparison is an important factor for monitoring student progress in reading comprehension from the second to the fourth grades. Finally, the use of two distinct measures including the exclusive use of one text genre enables psychologists and teachers to better diagnose reading comprehension difficulties and understand student performance in specific academic areas. With this information, it is possible to direct instruction to overcome the identified difficulties or develop intervention programs more focused on training strategies that address one or the other text genre.

This research was supported by Grant FCOMP-01-0114-FEDER-010733 from Fundação para a Ciência e Tecnologia (FCT) and the European Regional Development Fund (FEDER) through the European program COMPETE (Operational Program for Competitiveness Factors) under the National Strategic Reference Framework (QREN). It was also supported by FCT grant SFRH/BD/94763/2013.

REFERENCES


© 2016 Scandinavian Psychological Associations and John Wiley & Sons Ltd


Received 18 January 2016, accepted 23 August 2016