



Reading strategy use scale: an analysis using the rasch rating scale model

Irene Cadime¹ · Bruna Rodrigues¹ · Iolanda Ribeiro²

Accepted: 19 August 2022 / Published online: 15 November 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Although reading strategy use can be measured using different methods, self-report measures are particularly cost-effective in educational settings. The main goal of this study was to investigate the psychometric properties of the items of the Reading Strategy Use Scale (RSU), using the Rasch Rating Scale Model. Although the RSU has already been studied using classical analyses, some issues remain regarding the suitability of the response scale and the items' validity. A total of 179 students attending the fifth and sixth grades participated in this study. The results supported a unidimensional structure and local independence of the items. The seven-category response scale showed inadequate functioning. Rather, the results suggested that a five-category scale was more appropriate. Rasch reliability indicators were very high. There was no meaningful differential item functioning as a function of gender. A test of differences indicated that girls use reading strategies more frequently than boys. A positive correlation between the RSU scores and a reading comprehension measure was found. The findings support RSU as a robust self-reporting instrument to measure the frequency of use of reading strategies and highlight the usefulness of Rasch analysis in developing robust educational measures.

Keywords Reading strategy use · Rasch analysis · Differential item functioning · Gender

✉ Irene Cadime
ireneCadime@psi.uminho.pt

¹ Psychology Research Centre, School of Psychology, University of Minho, Campus of Gualtar, 4710-057 Braga, Portugal

² School of Psychology, University of Minho, Minho, Portugal

Introduction

Reading strategies consist of behaviors or thoughts intentionally activated by individuals to understand and attribute meaning to a text and achieve their reading goals (Afflerbach et al., 2008; Miyamoto et al., 2019). These include cognitive and metacognitive strategies that can be activated before, during and after reading a text (Mason, 2013). Cognitive strategies involve the construction of meaning from the integration of new information into existing knowledge schemas, allowing the reader to understand the material read (Afsharrad et al., 2017). Examples of cognitive reading strategies are paraphrasing, questioning, scanning textual information, predicting content from images, titles or context, activating prior knowledge, summarizing, underlining key words and sentences, underlining unfamiliar words, taking notes and consulting the dictionary (Mokhtari & Reichard, 2002; Parkinson & Dinsmore, 2018). Metacognitive strategies refer to the knowledge and self-regulation of cognitive processes, as well as the monitoring and assessment of the achieved level of comprehension (Erler & Finkbeiner, 2007). These allow the reader to plan the reading activity, monitor and adapt their cognitive activity while reading, analyze the effectiveness of the strategies used and readjust them (Afsharrad et al., 2017; Griva et al., 2012). Hence, several studies have consistently found, across a wide range of grade levels, a connection between the use of reading strategies and the reading comprehension levels attained (Ammel & Keer, 2021; Frid & Friesen, 2020; Köse & Güneş, 2021; Liao et al., 2022). Research has also indicated that programs centered on promoting the use of the aforementioned strategies have a positive impact on students' reading and language skills (Bippert, 2020; Mason, 2013). Therefore, the availability of robust measures to assess reading strategy use is of utmost importance for practice and for research, whether it is for purposes of identifying students performing poorly, monitoring progress or assessing the effects of intervention programs.

Reading strategy use has been assessed using different methods, which can be grouped into online and offline methods. The first group includes, among others, think-aloud protocols (e.g., Cromley & Wills 2016; Meyers et al., 1990; Wang, 2016) and eye movement registrations (e.g., Tremblay et al., 2021). In both cases, the behavior of the readers during a reading task is registered and coded. Although the scores obtained with these methods are the ones that are most highly correlated with reading comprehension (Cromley & Azevedo, 2006), they still have some disadvantages. Veenman (2011) points out weaknesses such as the inability of some individuals to verbally report what they are doing in think-aloud protocols, the fact that these methods can be intrusive for some people and, most of all, that they are very time-consuming and labor intensive. In the case of eye-movement registration methods, these also require costly equipment that may not be available or easily accessible. Offline methods are those in which reading strategy use is assessed before or after the reading task. One example of this type of method consists of requiring the use of a specific strategy during a reading task and asking comprehension questions afterward (Spörer et al., 2009). Although this is a method that provides some evidence on whether students are able to use strategies with efficacy, concerns have been raised that this is not a pure measure of reading strategy use, as the scores also measure comprehension itself (Muijselaar et al., 2017). Hence, the most com-

monly used off-line measures of reading strategy use are questionnaires. These can be divided into questionnaires that address the students' knowledge of reading strategies and questionnaires that address the frequency of use of reading strategies. In the first type, students are asked what they do in specific situations, such as what they do when they stop understanding the text (Muijselaar & de Jong, 2015). The second type of questionnaire is essentially self-reporting measures, where respondents indicate the frequency with which they use specific strategies during reading (Liao et al., 2022; Mokhtari & Reichard, 2002). Self-reporting questionnaires are the most commonly used methods (Cromley & Azevedo, 2006), despite having been occasionally criticized – for example, for being prone to social desirability and memory distortions (Veenman, 2011), for not providing information on how effectively readers use the strategies but merely on how often they use them (Miyamoto et al., 2019), or for being less related to reading measures than on-line measures (Cromley & Azevedo, 2006). Although these limitations should be considered, these instruments are still useful in various settings due to their cost-effectiveness relationship and convenience of use, as they can be easily administered to large groups and can be quickly scored (Gascoine et al., 2017; Veenman et al., 2006).

The Reading Strategy Use (RSU) scale is a self-reporting measure of the frequency of use of cognitive and metacognitive reading strategies that was originally developed in New Zealand to assess children aged approximately 10–13 years old (Pereira-Laird & Deane, 1997). A two-factor structure – with a cognitive and a metacognitive factor – was established through confirmatory factor analysis. Both factors had adequate reliability as measured by Cronbach's alpha (0.73 and 0.85 for the cognitive and metacognitive factors, respectively). Validity evidence was provided by correlations with scores in measures of reading comprehension, vocabulary, and study skills, as well as with the classroom marks. The RSU was later modified to be administered to younger children (attending the second grade) in the United States (Reutzell et al., 2005). Among the modifications was the shrinkage of the response scale from seven categories to three categories. However, the psychometric properties of this modified version were not explored. In 2015, the version for older children was adapted for European Portuguese (Ribeiro et al., 2015). Contrarily to the results of Pereira-Laird & Deane (1997), a one-factor structure obtained a better fit than a two-factor structure. Reliability was high ($\alpha=0.85$) and the RSU scores were positively correlated with school marks. Additionally, a test of differences indicated higher RSU scores reported by girls, compared to boys (Ribeiro et al., 2015). Although, studies of the RSU have shown adequate psychometric properties, the measure has been criticized (e.g., Mokhtari & Reichard 2002) for including items that do not seem to be reading strategies. In the study of adaptation for the Portuguese population (Ribeiro et al., 2015), one item that do not seem to be a good representative of the construct – “After I have been reading for a short time, the words stop making sense” –, was dropped, as it had a very low factor loading. More detailed analyses of the items can lead to the identification of additional items with validity issues and thus improve the measure. Additionally, the functioning of the response scale of the RSU has never been studied. Research has made an effort to provide guidelines for the optimal number of categories in Likert-type response scales, but has obtained variable results, suggesting a number varying between 4 and 10 (Lozano

et al., 2008; Preston & Colman, 2000). On the one hand, having a very low number of categories can lead to low response variability and having an excessively high number of categories can foster extreme responding, either way decreasing validity (Cox, 1980). Moreover, RSU has been used to test for gender comparisons (Ribeiro et al., 2015), but no evidence of measurement invariance has been provided. Measurement invariance is a fundamental property of any assessment instrument, as it assures that the performance of an individual on it depends only on their level in the latent variable and not on their group of origin. Therefore, fair comparisons between groups should be based on nonbiased items. Differential item functioning (DIF) analyses are a statistical procedure applied at the item level to check whether the items measure aspects other than the latent variable (Walker, 2011). The presence of DIF suggests the presence of bias in the item. Thus, the assertion of the existence of gender differences can only be made after guaranteeing the absence of DIF in the instrument used to measure reading strategy use.

The goal of this study is to address the psychometric properties of the items of the European Portuguese version of the RSU using Rasch model analyses. These analyses have several advantages compared to the most traditional ones in the scope of classical test theory. Among the most relevant are the properties of conjoint measurement and specific objectivity (de Ayala, 2009). In Rasch model analysis, two parameters are estimated: an item parameter (b_i) for each individual item, traditionally designated item difficulty, and an ability parameter for each person (θ_n). These are estimated conjointly and placed on a common logit (logarithm of odds) scale, thus being of interval level. These estimates can be visualized in a continuum on which the items and persons are ordered according to their respective parameter values. Thus, “because of the interval level data and conjoint measurement scale between person ability and item difficulty, Rasch measurement indices are considered item and sample independent” (Sondergeld & Johnson, 2014, p. 584). This property is known as specific objectivity and establishes that the difference between two people in a skill should not depend on the specific items with which it was estimated, and that the difference between two items should not depend on the subjects used to estimate its parameters. Thus, if the data effectively fit the Rasch model, the comparisons between people will be independent of the administered items and the estimates of the parameters of the items will not be influenced by the distribution of the sample that is used (Prieto & Delgado, 2003). Moreover, Rasch analysis provides indicators of how well each item fits the underlying construct (Bond & Fox, 2007), which can be useful to flag RSU items that may not be measuring adequately reading strategy use. Although the Rasch model was originally developed to deal with dichotomous data, the model was extended to polytomous data – the Rasch Rating Scale Model (RSM; Andrich 1978). The RSM has an additional advantage: it allows testing of the functioning of the categories of the Likert scales (Bond & Fox, 2007). Moreover, although there are several statistical procedures to examine the presence of DIF, Rasch-based DIF analyses are one of the most commonly used methods to detect gender-based DIF in educational assessments (Aryadoust et al., 2012). Hence, the specific goals of this study were (a) to provide additional evidence of unidimensionality for the RSU; (b) to assess reliability and the fit of the RSU items to the Rasch RSM; (c) to examine the functioning of the seven-category response scale of the

RSU; (d) to investigate the existence of gender-based DIF in the RSU scores; and (e) to explore the relationship between the RSU scores and reading comprehension.

Regarding the first goal, we expect to obtain evidence of unidimensionality, as this structure already obtained the best fit in the previous study with the Portuguese RSU versions (Ribeiro et al., 2015) and unidimensionality is a pre-requisite for the estimation of Rasch model. As the item fit and the response scale of the RSU were never explored, no predictions are presented. Regarding the fourth goal, we expect that the RSU items show no meaningful gender-based DIF, so that fair comparisons between boys and girls in the scores can be made. As research has found some gender differences, with girls reporting using more reading strategies than boys (Afsharrad et al., 2017; Griva et al., 2012; Köse & Güneş, 2021), we expect to replicate this result with RSU, after guaranteeing the absence of meaningful gender-related DIF. Regarding the last goal, we expect to find a positive relation between the RSU scores and the scores in a reading comprehension measure, as previous research has clearly demonstrated a positive correlation between reading strategy use and comprehension, i.e., readers who use reading strategies are likely to achieve better reading comprehension (Follmer & Sperling, 2018; Köse & Güneş, 2021; Liao et al., 2022).

Method

Participants and procedures

The sample was retrieved from a study on reading comprehension [reference omitted]. The study was authorized by the ethics committee of the University of Minho. As the data were collected in schools, authorizations of the school boards and of the Portuguese Ministry of Education were also collected. Written informed consent was also collected from students' parents or other legal tutors. The participants were administered a battery of tests that included the RSU and reading measures. The RSU and the reading comprehension test were administered to students in group (class) in the students' classrooms by psychologists with experience in administering these measures. Each measure was administered in two different sessions. Students took about 10–15 min to complete RSU and about 45 min to complete the reading comprehension measure.

A total of 179 students participated, of whom 94 (52.5%) attended fifth grade (mean age = 11; $SD=0.55$) and 85 (47.5%) attended sixth grade (mean age = 12; $SD=0.50$) in public schools in northern Portugal. The number of boys ($N=90$) and girls ($N=89$) was similar. Approximately 45% of the students ($N=80$) benefited from school social support, which was provided to students from low socioeconomic levels. Most of the children's mothers had completed only elementary education (59%), whereas 25.9% completed secondary education and 15.1% had a higher education degree.

Measures

Reading Strategy Use (RSU) Scale (Pereira-Laird & Deane, 1997). This is a self-report instrument composed of 22 items that measure students' frequency of use of cognitive and metacognitive reading strategies when reading narrative and expository texts. Each item consists of a statement that represents a reading strategy, and students should indicate how frequently they use it using a seven-category scale with the following descriptors: 1 = *never*; 2 = *almost never*, 3 = *seldom*, 4 = *sometimes*, 5 = *often*, 6 = *almost always*, 7 = *always*. Although a two-factor structure had the best fit in New Zealand (Pereira-Laird & Deane, 1997), the results of the validation study for European Portuguese (Ribeiro et al., 2015) suggested a one-factor structure and a high reliability (Cronbach's $\alpha=0.85$). Evidence of validity was also provided, as indicated by significant correlations between the RSU scores and students' school scores. The referenced study (Ribeiro et al., 2015) also suggested that one item should be dropped, and therefore, in this study, a version with 21 items was used. The items can be found in the [Appendix](#).

Test of Reading Comprehension of Narrative Texts (TRC-n; Rodrigues et al., 2020; Santos et al., 2016). This test is composed of five vertically scaled test forms that assess reading comprehension of narrative texts in students from grades 2 to 6. In the present study, the test forms for grades 5 and 6 (TRC-n-5 and TRC-n-6, respectively) were administered. Students should read silently a group of texts and the respective multiple-choice questions (three options) and to mark their responses on an answer sheet. The responses are scored as 0 (incorrect) or 1 (correct). The total raw score of each test form is converted to a standardized score that is placed in a common metric for all test forms. The reliability coefficients are high for the test forms for grades 5–6, ranging between 0.72 and 0.97 (Rodrigues et al., 2020). Regarding validity evidence, scores obtained in these test forms are statistically correlated with scores in other measures of language and reading (Rodrigues et al., 2020).

Data analysis

Rasch RSM analyses were carried out using Winsteps Version 3.61.1 (Linacre & Wright, 2001). To check for unidimensionality, principal component analysis of the Rasch standardized residuals (PCAR) was carried out (Chou & Wang, 2010). The residual is the difference between an observed response and that predicted by the model. To determine the presence of a dominant dimension in the data, two requirements must be met: (a) the percentage of the variance explained by the main dimension must be at least 20% and (b) the eigenvalue of the first secondary dimension must be lower than 3 (Miguel et al., 2013). Correlations of the residuals were also computed to check the local independence of the items, i.e., the measurement principle that asserts that the score in an item depends solely on the person's latent trait and is not influenced by the responses to other items. Coefficients lower than 0.70 provide evidence for the local independence of the items (Linacre, 2011). To assess the functioning of the response scale, the following requirements were examined: (a) a monotonic increase of the average measures, which indicates that persons presenting a high level on the latent trait will endorse higher response categories; (b) the

nonexistence of a response category with an infit or outfit higher than 2.00; and (c) a monotonic increase in the step calibrations (Bond & Fox, 2007; Linacre, 2002a). The step calibrations are the *logit* values where categories k and $k-1$ have the same probability of being endorsed. Ordered steps indicate that each category is the most likely to be observed at certain intervals of the measurement scale. Item fit was assessed by analyzing the mean square (MNSQ) infit and outfit statistics. Values between 0.5 and 1.5 indicate a good fit (Gómez et al., 2012). Moreover, the values should not be higher than 2.0 because they suggest severe misfit (Linacre, 2002b). Item-total correlations were also examined, and a minimum of 0.40 was stipulated. Reliability was checked by means of the item separation reliability (ISR) and person separation reliability (PSR) indices. ISR is an estimate of how likely it is to achieve the same ranking of the items in the measured variable given a different sample of comparable ability, and PSR is an estimate of how likely it is to achieve the same ordering of the persons if those were given another set of items that measured the same construct (Bond & Fox, 2007; Gómez et al., 2012). Both reliability estimates range between 0 and 1, and a minimum of 0.70 is advisable. Finally, the presence of gender-based DIF was examined, with boys being the reference group and girls the focal group. The Rasch model DIF statistics are based on the comparison of the difficulty parameters of each item obtained by each group. Empirical evidence of the presence of DIF is provided by statistically significant ($p < .05$) results in Rasch-Welch's t test. However, the impact of uniform DIF on test scores should be considered meaningful only when it is found in more than 25% of the items within a dimension (Rouquette et al., 2019) and DIF contrast is higher than 0.50 logit (Linacre, 2011). After checking the absence of DIF, a comparison of differences between boys and girls in mean estimates was performed using Welch's t test. The Pearson correlation coefficient and linear regression were used to analyze the relationship between the RSU scores and the scores in the measure of reading comprehension.

Results

Dimensionality and local independence

The results of the PCAR indicated that the variance explained by the measures was 57.8%. The unexplained variance of the first contrast of residuals was 4.5% and had an eigenvalue of 2.2. These values provide evidence for the unidimensionality of the measure. The largest standardized residual's correlation was 0.31, suggesting the absence of local dependence of the items.

Item parameters and category statistics

Table 1 presents the descriptive statistics for the raw scores in each item. For most items there was a balanced distribution of the responses in the seven categories. However, there were some items more prone to extreme responses. This was the case particularly for the two reverse-coded items – Item 14 (“When some of the sentences

Table 1 Descriptive statistics of the items' raw scores

Item	Scores							
	M (SD)	1	2	3	4	5	6	7
		n (%)	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
RSU_1	4.07 (1.82)	22 (12.3)	18 (10.1)	22 (12.3)	43 (24.0)	27 (15.1)	31 (17.3)	16 (8.9)
RSU_2	4.63 (1.50)	7 (3.9)	6 (3.4)	22 (12.3)	51 (28.5)	41 (22.9)	30 (16.8)	22 (12.3)
RSU_3	5.71 (1.62)	5 (2.8)	5 (2.8)	8 (4.5)	25 (14.0)	18 (10.1)	33 (18.4)	85 (47.5)
RSU_4	2.69 (1.77)	70 (39.1)	25 (14.0)	26 (14.5)	32 (17.9)	6 (3.4)	15 (8.4)	5 (2.8)
RSU_5	5.03 (1.50)	3 (1.7)	9 (5.0)	14 (7.8)	40 (22.3)	32 (17.9)	50 (27.9)	31 (17.3)
RSU_6	4.09 (1.85)	19 (10.6)	25 (14.0)	22 (12.3)	36 (20.1)	28 (15.6)	30 (16.8)	19 (10.6)
RSU_7	3.80 (1.93)	28 (15.6)	27 (15.1)	25 (14.0)	35 (19.6)	18 (10.1)	29 (16.2)	17 (9.5)
RSU_8	4.84 (1.76)	15 (8.4)	4 (2.2)	19 (10.6)	29 (16.2)	34 (19.0)	46 (25.7)	32 (17.9)
RSU_9	3.96 (1.75)	19 (10.6)	23 (12.8)	25 (14.0)	49 (27.4)	17 (9.5)	35 (19.6)	11 (6.1)
RSU_10	4.80 (1.65)	9 (5.0)	8 (4.5)	19 (10.6)	37 (20.7)	36 (20.1)	40 (22.3)	30 (16.8)
RSU_11	3.72 (1.84)	32 (17.9)	20 (11.2)	27 (15.1)	38 (21.2)	25 (14.0)	24 (13.4)	13 (7.3)
RSU_12	4.47 (1.76)	12 (6.7)	16 (8.9)	18 (10.1)	49 (27.4)	22 (12.3)	38 (21.2)	23 (13.4)
RSU_13	5.46 (1.82)	10 (5.6)	8 (4.5)	8 (4.5)	24 (13.4)	18 (10.1)	35 (19.6)	76 (42.5)
RSU_14	1.73 (1.28)	122 (68.2)	22 (12.3)	9 (5.0)	19 (10.6)	3 (1.7)	3 (1.7)	1 (0.6)
RSU_15	5.07 (1.79)	7 (3.9)	13 (7.3)	14 (7.8)	35 (19.6)	22 (12.3)	34 (19.0)	54 (30.2)
RSU_16	3.06 (1.91)	56 (31.3)	28 (15.6)	21 (11.7)	33 (18.4)	16 (8.9)	14 (7.8)	11 (6.1)
RSU_17	4.67 (1.66)	13 (7.3)	7 (3.9)	14 (7.8)	45 (25.1)	37 (20.7)	39 (21.8)	24 (13.4)
RSU_18	4.03 (1.88)	22 (12.3)	25 (14.0)	19 (10.6)	41 (22.9)	23 (12.8)	30 (16.8)	19 (10.6)
RSU_19	3.49 (2.04)	41 (22.9)	29 (16.2)	25 (14.0)	30 (16.8)	14 (7.8)	19 (10.6)	21 (11.7)
RSU_20	5.41 (1.67)	3 (1.7)	13 (7.3)	10 (5.6)	25 (14.0)	20 (11.2)	46 (25.7)	62 (34.6)
RSU_21	4.87 (1.65)	10 (5.6)	6 (3.4)	16 (8.9)	40 (22.3)	33 (18.4)	42 (23.5)	32 (17.9)

Note: for the reverse-coded items (RSU_14 and RSU_16), these statistics refer to the scores before recoding

that I am reading are hard, I give up the reading”) and Item 16 (“When I cannot read a word in the story, I skip it”) –, as they had a large portion of the responses concentrated in the lower end of the scale. In contrast, items 3, 13 and 20 were highly endorsed, suggesting that the strategies described in these items (slow-down reading, go-back in the text and re-reading) are quite frequently used by students.

Regarding the items' Rasch RSM parameters, the two reverse-coded items (items 14 and 16) had item-total correlations lower than 0.40. Item 16 also had an MNSQ outfit higher than 2.0. Therefore, these two items were deleted, and the data were again calibrated. After deleting these items, all item-total correlations were higher than 0.40, and no item had an MNSQ infit or outfit higher than 2.0. Table 2 presents the category statistics for this calibration. The seven-category scale presented a regular distribution of the category frequencies, a monotonic increase in the average measures, and all response categories had fit statistics below 2.00. However, there was no monotonic increase in the step calibrations, as the steps/thresholds were disordered.

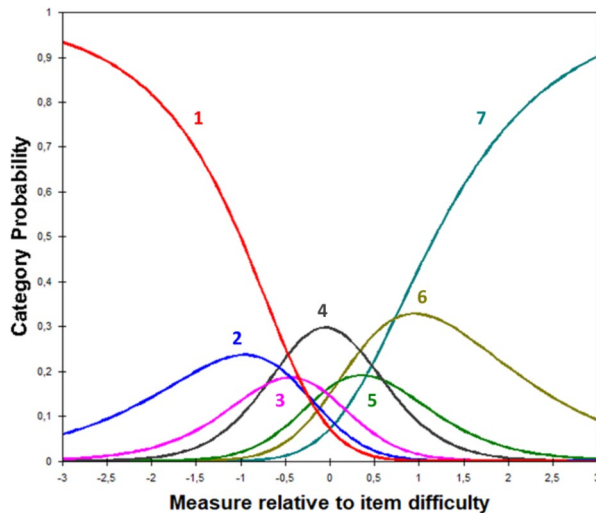
Figure 1 presents the category probability curves for the seven-category response scale. Each category should have a distinct peak in the probability curve graph, thus indicating that each category is indeed the most probable for some portion of the measured variable (Bond & Fox, 2007). As shown in Fig. 1, categories two, three and five were never the most likely ones in the continuum. These results suggested that collapsing categories can improve the efficiency of the response scale. The data

Table 2 Category statistics and reliability

Likert scale	Observed count	Observed average	MNSQ		Step	Category estimate	PSR	ISR
			Infit	Outfit				
Seven-category								
1	347	-0.59	0.95	0.99	-----	-1.94		
2	287	-0.28	1.09	1.20	-0.26	-0.95		
3	353	-0.10	1.02	1.07	-0.44	-0.44	0.89	0.98
4	704	0.06	0.99	1.01	-0.71	-0.05		
5	471	0.32	0.85	0.86	0.59	0.36		
6	646	0.50	1.05	1.06	0.09	0.95		
7	593	0.77	1.04	1.04	0.73	2.16		
Five- category								
1	347	-0.90	0.95	0.98	-----	-2.53		
2	640	-0.28	1.08	1.15	-1.21	-1.02		
3	704	0.09	0.98	1.00	-0.20	-0.07	0.89	0.98
4	1117	0.62	0.94	0.94	-0.10	0.98		
5	593	1.13	1.06	1.05	1.50	2.74		

Note. PSR=person separation reliability; ISR=item separation reliability; MNSQ=mean square

Fig. 1 Category probability curves for the seven-category response scale



were calibrated again, and the response scale was recoded to a five-category scale. To maintain the coherence of the scale, we collapsed categories 2 (*almost never*) and 3 (*seldom*), as well as categories 5 (*often*) and 6 (*almost always*).

The results for the five-category response scale indicated that the steps increased monotonically, each of the categories was the most likely one across the continuum, the fit statistics improved slightly and there was no decrease in reliability (see Table 2; Fig. 2). Regarding reliability, both PSR and ISR were very high (see Table 2). Table 3 presents the item statistics for this last calibration after recoding the response scale.

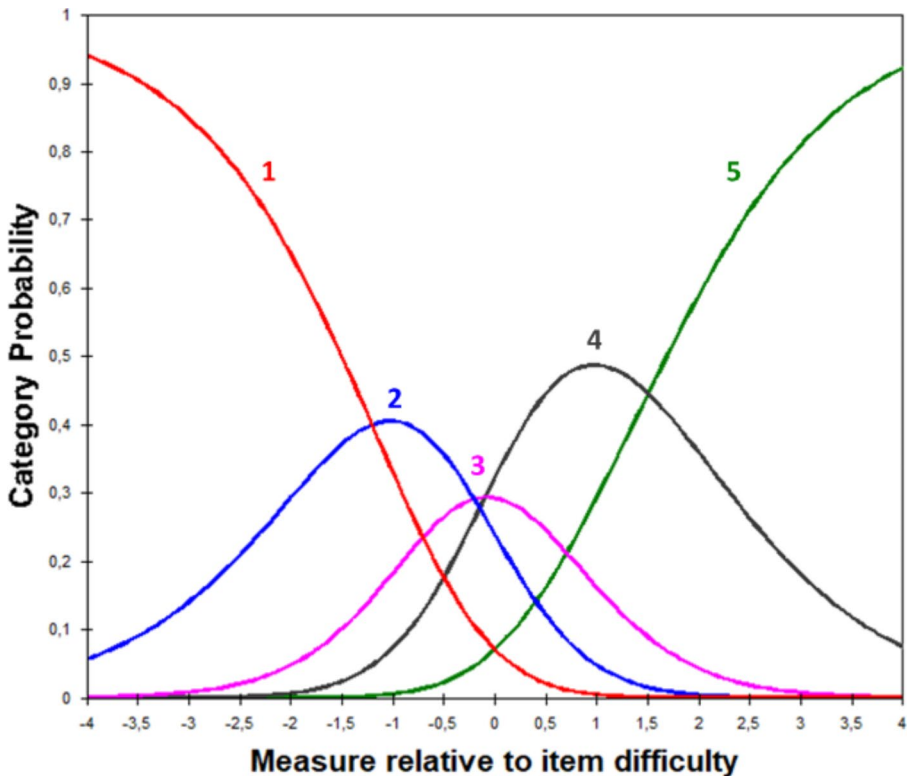


Fig. 2 Category probability curves for the five-category response scale

All items presented an item-total correlation higher than 0.40 and infit and outfit lower than 2.00, suggesting an adequate fit.

Differential item functioning (DIF) and relationship with comprehension

Table 4 presents the results of DIF analyses. Four items (Items 9, 10, 13 and 20) had statistically significant DIF. Girls had higher average estimates than boys in Items 9 and 10, and the opposite was verified for Items 13 and 20, in which boys had higher average estimates. However, the size of the contrast was lower than 0.50 in all four cases, and therefore, the DIF is not meaningful. When comparing mean estimates for the 19 final items, girls (mean estimate=0.40, $SD=0.74$) obtained significantly higher scores than boys (mean estimate=0.15, $SD=0.75$), $t(176) = -2.245$, $p = .026$.

A significant and positive relationship was found between the total scores of this revised version of RSU and the scores in reading comprehension ($r = .246$; $p < .001$). The RSU scores explained about 6% of the variance in reading comprehension ($\beta = 0.246$; $R^2 = 0.06$)

Table 3 Reading Strategy Use Scale: item statistics

Item	Estimate (SE)	Item-total correlation	MNSQ	
			Infit	Outfit
RSU_1	0.29	0.60	0.94	0.91
RSU_2	-0.11	0.50	0.88	0.95
RSU_3	-1.02	0.42	1.53	1.50
RSU_4	1.31	0.59	1.03	0.97
RSU_5	-0.38	0.52	0.82	0.88
RSU_6	0.26	0.58	1.00	0.98
RSU_7	0.48	0.57	1.06	1.13
RSU_8	-0.23	0.58	1.04	1.01
RSU_9	0.40	0.54	0.91	0.92
RSU_10	-0.23	0.58	0.90	0.89
RSU_11	0.54	0.66	0.87	0.91
RSU_12	0.03	0.63	0.81	0.80
RSU_13	-0.78	0.43	1.60	1.68
RSU_15	-0.44	0.51	1.20	1.24
RSU_17	-0.14	0.56	0.89	0.89
RSU_18	0.31	0.71	0.75	0.75
RSU_19	0.68	0.50	1.37	1.44
RSU_20	-0.72	0.59	1.01	0.98
RSU_21	-0.26	0.66	0.76	0.73

Note: SE=standard error;
MNSQ=mean square;
Estimate= β_i

Table 4 Results of DIF analyses as a function of gender

Item	Estimate		DIF contrast	Rasch-Welch		
	Boys	Girls		<i>t</i>	df	<i>p</i>
RSU_1	0.28	0.30	0.02	0.14	177	0.889
RSU_2	-0.13	-0.09	0.04	0.29	177	0.776
RSU_3	-0.90	-1.18	-0.28	-1.45	177	0.149
RSU_4	1.34	1.28	-0.06	-0.34	177	0.733
RSU_5	-0.32	-0.45	-0.13	-0.81	177	0.419
RSU_6	0.27	0.25	-0.01	-0.09	177	0.926
RSU_7	0.52	0.44	-0.08	-0.54	177	0.592
RSU_8	-0.27	-0.18	0.09	0.57	177	0.573
RSU_9	0.19	0.60	0.42	2.74	177	0.007
RSU_10	-0.42	-0.02	0.40	2.48	177	0.014
RSU_11	0.59	0.48	-0.12	-0.77	177	0.444
RSU_12	0.14	-0.09	-0.23	-1.46	177	0.146
RSU_13	-0.60	-0.99	-0.40	-2.18	177	0.030
RSU_15	-0.54	-0.32	0.22	1.31	177	0.191
RSU_17	-0.21	-0.06	0.14	0.91	177	0.364
RSU_18	0.19	0.43	0.24	1.59	177	0.113
RSU_19	0.81	0.56	-0.25	-1.63	177	0.105
RSU_20	-0.53	-0.94	-0.41	-2.29	177	0.023
RSU_21	-0.33	-0.19	0.14	0.87	177	0.383

Note. Df=degrees of freedom; DIF=Differential Item Functioning

Discussion

The main goal of this study was to explore the psychometric properties of the items of the European Portuguese version of the RSU using Rasch RSM model analyses. The first specific goal was to provide evidence of unidimensionality for the RSU. In addition to being a requirement for the Rasch model, a one-dimensional structure had already been previously found in a sample of Portuguese students using confirmatory factor analysis (Ribeiro et al., 2015). The results of this study show that unidimensionality is replicable with a different method to investigate the internal structure of the test. This factor structure contrasts with the one obtained in the study of the original version in New Zealand, where a two-factor structure, distinguishing cognitive and metacognitive strategies, had the best fit (Pereira-Laird & Deane, 1997). In fact, there has been some debate on the differentiation between cognitive and metacognitive strategies, as they are interdependent and sometimes difficult to distinguish. As Veenman (2011) states, “higher-order metacognitive processes monitor and regulate lower-order cognitive processes that, in turn, shape behavior. Thus, drawing inferences is a cognitive activity, but the self-induced decision to initiate such activity is a metacognitive one” (p. 205). In addition, research suggest that cognitive strategy use may not be highly effective without the concomitant use of metacognitive strategies (Zhao et al., 2014). For example, making an outline of what is being read is effective mainly when the reader is able to monitor the process, assess it and revise it when needed. Another example: underlining the main ideas or key concepts involves distinguishing between those that are relevant and those that are not, and this process requires a constant monitoring as the reading advances, and sometimes going back and forth in the text to revise the underlined information. Other studies suggest that promoting the use of cognitive strategies, such as the use of concept mapping, leads to an improvement of metacognitive skills (e.g., Welter et al., 2022). Thus, the one-factor structure observed in the measure analyzed in this study may reflect the fact that cognitive and metacognitive strategies are probably mobilized together frequently.

A second goal was to assess reliability and the fit of the RSU items to the Rasch RSM. The results suggested a high reliability, and all items had a good fit to the model. However, two items were discarded due to low item-total correlations. These items were reverse-coded. Although the inclusion of reverse-coded items has been a long-term recommended practice to avoid the acquiescence effect in self-report measures (e.g., Nunnally 1978), some more recent research has discouraged it. A considerable number of studies have shown that reverse-coded items decrease model fit and frequently form separate dimensions that lack meaningfulness, which is particularly problematic in unidimensional tests (Cassady & Finch, 2014; Clauss & Bardeen, 2020; Essau et al., 2012; Woods, 2006). The reason for this finding may be related to some evidence that suggests that reverse-coded items and straightforward items may involve different cognitive processes and thus not measure the same latent trait (Suárez-Alvarez et al., 2018; Weems & Onwuegbuzie, 2001). Additionally, research has also suggested that combining both types of items in the same test decreases the variability in the responses and leads to worse discriminative power and reliability (Suárez-Alvarez et al., 2018; Vigil-Colet et al., 2020). A content analysis showed

additional problems, particularly in the case of Item 16. The adequacy of the strategy included in this item – “When I cannot read a word in the story, I skip it” – is highly ambiguous. On the one hand, skipping difficult words or text parts seems to be a strategy much more used by poor readers than by good readers (Anastasiou & Griva, 2009). Moreover, contrary to poor readers, good readers focus more on constructing the meaning of the text as a whole, instead of focusing on understanding all single words (Lau, 2006), and therefore prefer to use strategies such as activating previous knowledge or imagery (Anastasiou & Griva, 2009; Lau & Chan, 2003). We should also note that skipping difficult words can sometimes be an adequate reading strategy: in some cases, that meaning can be inferred later, as the reading advances further in the text, and, in other cases, not knowing the meaning of some specific words simply does not hinder the comprehension of the text as a whole (Giasson, 2000). For all these reasons, we recommend that Items 14 and 16 should be dropped from RSU.

The third goal of the study was to examine the functioning of the seven-category response scale of the RSU. This analysis is one of the potentialities of the Rasch RSM model (Bond & Fox, 2007). The results indicated that the seven-category response scale was inadequate, as some categories were redundant, i.e., they were never the most likely to be endorsed. A five-category response scale showed better results. The functioning of the response scale in reading strategy use questionnaires has seldom been investigated. However, a five-category Likert scale, ranging from *never* to *always*, is frequently the option – see, for example, one of the most used measures, the Metacognitive Awareness of Reading Strategies Inventory (MARSII; Mokhtari & Reichard 2002). The results of our study provide empirical evidence for the adequacy of this option.

The fourth specific goal was to investigate the existence of gender-based DIF in the scores of the RSU, as the presence of DIF means that any inferences made from the scores are necessarily biased. The results of our study suggest that although four items obtained statistically significant DIF, the effect size was negligible. Therefore, fair comparisons between boys and girls can be made in reading strategy use assessed with the RSU. Consistent with previous research (Afsharrad et al., 2017; Griva et al., 2012; Köse & Güneş, 2021), the results of our study indicate that girls use reading strategies more frequently than boys.

The final goal was to explore the relationship between the RSU scores and the scores in a reading comprehension measure. Given that previous research has clearly demonstrated a correlation between reading strategy use and reading comprehension (Follmer & Sperling, 2018; Köse & Güneş, 2021; Liao et al., 2022), we expected to find a positive correlation between the scores of both measures, which was indeed found in the results of our study. Nonetheless, the relationship was weak, with reading strategy use explaining only 6% of the variance observed in reading comprehension. Studies in European Portuguese with children in grades 4 to 6 show that, at this stage, reading comprehension still depends heavily on basic reading skills, such as oral reading fluency, and on linguistic skills, such as vocabulary (Fernandes et al., 2017; Rodrigues et al., 2022). Thus, reading comprehension performance can be explained to a lesser extent by reading strategy use. Hence, the findings of this study provide evidence of validity for the revised version of RSU with 19 items and a five-category response scale.

Conclusion

Overall, the findings of the present study add evidence of reliability and validity for the Portuguese version of the RSU, confirming it as a robust measure to assess the frequency of use of reading strategies in students from the fifth and sixth grades. Rasch RSM modeling established the unidimensionality of the scale, suggested changes to be introduced in the Likert scale, and provided evidence of the nonexistence of gender bias. These findings suggest that measures can be improved when a more detailed examination of the items and the response scale are performed, and Rasch modeling offers several possibilities for these analyses that complement the more traditional ones, such as factor analyses. The findings of our study can also serve as a reference framework for the study of the psychometric properties of the modified version of the RSU for younger children (Reutzel et al., 2005), which are yet to be explored. The RSU is a relatively short instrument and can be administered in large groups, making it especially useful for use in educational settings or in research when large amounts of data must be collected in short amounts of time.

Appendix

Items in Portuguese and English.

Item	Portuguese	English
RSU_1	Antes de ler uma passagem do texto, faço uma leitura rápida para ficar com uma ideia global.	I read quickly through the whole passage to get the general idea before I read it thoroughly.
RSU_2	Aprendo novas palavras, relacionando-as com palavras que já conheço.	I learn new words by relating/linking them with words which I already know.
RSU_3	Quando um capítulo do meu livro é difícil de entender, leio mais devagar.	When I find that a chapter in my book is hard to understand, I slow down my reading.
RSU_4	Faço um esquema do que estou a ler.	I make an outline of what I am reading.
RSU_5	Quando leio, consigo decidir que informações são mais ou menos importantes.	I am able to decide between more important and less important information while reading.
RSU_6	Quando estou a ler, às vezes paro para rever o que já li.	When I'm reading, I stop once in a while and go over what I have read.
RSU_7	Para me ajudar a perceber o que li, digo o texto pelas minhas próprias palavras.	To help me understand what I have read, I say it in my own words.
RSU_8	Identifico se um texto é ou não difícil e, em função disso, ajusto a minha velocidade de leitura.	I decide how difficult my reading passage is and then adjust the speed of my reading accordingly.
RSU_9	Durante a leitura, decoro palavras e conceitos difíceis, apesar de não os compreender.	When reading, I learn by heart difficult words and ideas without understanding them.
RSU_10	Aprendo novas palavras, imaginando uma situação em que elas ocorrem.	I learn new words by picturing in my mind a situation in which they occur.
RSU_11	Às vezes paro a leitura e faço perguntas a mim mesmo para avaliar até que ponto entendo o que estou a ler.	I stop once in a while and ask myself questions to see how well I understand what I am reading.

Item	Portuguese	English
RSU_12	Depois de ler algo, sento-me e fico a pensar no que li para ver se percebi.	After reading something, I sit and think about it for a while to check my understanding.
RSU_13	Quando me perco a ler, regresso ao ponto onde comecei a ter problemas.	When I get lost while reading, I go back to the point where I first had trouble.
RSU_14 (R)	Quando leio frases que não compreendo desisto da sua leitura.	When some of the sentences that I am reading are hard, I give up the reading.
RSU_15	Quando leio, formo imagens mentais daquilo que estou a tentar compreender.	When I read, I form pictures in my mind of the things I am trying to understand.
RSU_16 (R)	Quando não compreendo uma palavra num texto, passo-a à frente e prossigo com a leitura.	When I cannot read a word in the story, I skip it.
RSU_17	Ao ler algo, tento ligar o que leio àquilo que já sei.	When reading about something I try to link it to what I already know.
RSU_18	Leio de forma crítica e reflexiva, ou seja, enquanto leio algo, avalio o que estou a ler.	I read critically or thoughtfully, that is while reading something, I judge what I am reading.
RSU_19	Quando estou a ler, sublinho as ideias principais.	When I read, I underline the main ideas.
RSU_20	Quando estou a ler e me apercebo de que não entendi bem alguma coisa, releio para tentar compreender.	When I find I do not understand something while reading, I read it again and try to figure it out.
RSU_21	Enquanto leio, verifico se estou a entender o significado da história, perguntando a mim próprio se as minhas ideias encaixam com a restante informação da história.	When reading, I check how well I understand the meaning of the story by asking myself whether the ideas fit with the other information in the story.

Note. (R)=reverse coded items

Funding This study was conducted at the Psychology Research Center of the School of Psychology, University of Minho, supported by the Portuguese Foundation for Science and Technology through the Portuguese State Budget (UIDB/PSI/01662/2020). The first and second authors were also supported by the Portuguese Foundation for Science and Technology (CEECIND/00408/2018 and SFRH/BD/129582/2017, respectively).

Data availability The dataset is available from the corresponding author on reasonable request.

References

- Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, 61(5), 364–373. <https://doi.org/10.1598/RT.61.5.1>
- Afsharrad, M., Reza, A., & Benis, S. (2017). Differences between monolinguals and bilinguals/ males and females in English reading comprehension and reading strategy use. *International Journal of Bilingual Education and Bilingualism*, 20(1), 34–51. <https://doi.org/10.1080/13670050.2015.1037238>
- Anastasiou, D., & Griva, E. (2009). Awareness of reading strategy use and reading comprehension among poor and good readers. *Ilkogretim Online*, 8(2), 283–297
- Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–573
- Aryadoust, V., Goh, C. C. M., & Kim, L. O. (2012). An investigation of differential item functioning in the MELAB listening test. *Language Assessment Quarterly*, 8(4), 361–385. <https://doi.org/10.1080/15434303.2011.628632>
- Bippert, K. (2020). Text engagement & reading strategy use: A case study of four early adolescent students. *Reading Psychology*, 41(5), 434–460. <https://doi.org/10.1080/02702711.2020.1768987>
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Lawrence Erlbaum

- Cassady, J. C., & Finch, W. H. (2014). Confirming the factor structure of the Cognitive Test Anxiety Scale: Comparing the utility of three solutions. *Educational Assessment, 19*, 229–242. <https://doi.org/10.1080/10627197.2014.934604>
- Chou, Y., & Wang, W. (2010). Checking dimensionality in Item Response models with principal component analysis on standardized residuals. *Educational and Psychological Measurement, 70*(5), 717–731. <https://doi.org/10.1177/0013164410379322>
- Clauss, K., & Bardeen, J. R. (2020). Addressing psychometric limitations of the Attentional Control Scale via bifactor modeling and item modification. *Journal of Personality Assessment, 102*(3), 415–427. <https://doi.org/10.1080/00223891.2018.1521417>
- Cox, E. P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research, 17*(4), 407–422. <https://doi.org/10.1177/002224378001700401>
- Cromley, J. G., & Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning, 1*, 229–247. <https://doi.org/10.1007/s11409-006-9002-5>
- Cromley, J. G., & Wills, T. W. (2016). Flexible strategy use by students who learn much versus little from text: transitions within think-aloud protocols. *Journal of Research in Reading, 39*(1), 50–71. <https://doi.org/10.1111/1467-9817.12026>
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. The Guilford Press
- Erler, L., & Finkbeiner, C. H. (2007). A review of reading strategies: Focus on the impact of first language. In A. D. Cohen, & E. Macaro (Eds.), *Language Learner Strategies: 30 years of Research and Practice* (pp. 187–206). OUP
- Essau, C. A., Olaya, B., Anastassiou-hadjicharalambous, X., Pauli, G., Gilvarry, C., Bray, D., Callaghan, J. O., & Ollendick, T. H. (2012). Psychometric properties of the Strength and Difficulties Questionnaire from five European countries. *International Journal of Methods in Psychiatric Research, 21*(3), 232–245. <https://doi.org/10.1002/mp.1364>
- Fernandes, S., Querido, L., Verhaeghe, A., Marques, C., & Araújo, L. (2017). Reading development in European Portuguese: relationships between oral reading fluency, vocabulary and reading comprehension. *Reading and Writing, 30*(9), 1987–2007. <https://doi.org/10.1007/s11145-017-9763-z>
- Follmer, D. J., & Sperling, R. A. (2018). Interactions between reader and text: Contributions of cognitive processes, strategy use, and text cohesion to comprehension of expository science text. *Learning and Individual Differences, 67*, 177–187. <https://doi.org/10.1016/j.lindif.2018.08.005>
- Frid, B., & Friesen, D. C. (2020). Reading comprehension and strategy use in fourth- and fifth-grade French immersion students. *Reading and Writing, 33*(5), 1213–1233. <https://doi.org/10.1007/s11145-019-10004-5>
- Gascoine, L., Higgins, S., & Wall, K. (2017). The assessment of metacognition in children aged 4–16 years: a systematic review. *Review of Education, 5*(1), 58–59. <https://doi.org/10.1002/rev3.3079>
- Giasson, J. (2000). *A compreensão na leitura [Reading comprehension]* (2nd ed.). Edições Asa
- Gómez, L. E., Arias, B., Verdugo, M., & Navas, P. (2012). Application of the Rasch rating scale model to the assessment of quality of life of persons with intellectual disability. *Journal of Intellectual & Developmental Disability, 37*(2), 141–150. <https://doi.org/10.3109/13668250.2012.682647>
- Griva, E., Alevriadou, A., & Semoglou, K. (2012). Reading preferences and strategies employed by primary school students: Gender, socio-cognitive and citizenship issues. *International Education Studies, 5*(2), 24–34. <https://doi.org/10.5539/ies.v5n2p24>
- Köse, N., & Güneş, F. (2021). Undergraduate students' use of metacognitive strategies while reading and the relationship between strategy use and reading comprehension skills. *Journal of Education and Learning, 10*(2), 99. <https://doi.org/10.5539/jel.v10n2p99>
- Lau, K., & Chan, D. W. (2003). Reading strategy use and motivation among Chinese good and poor readers in Hong Kong. *Journal of Research in Reading, 26*(2), 177–190. <https://doi.org/10.1111/1467-9817.00195>
- Lau, K. (2006). Reading strategy use between Chinese good and poor readers: a think-aloud study. *Journal of Research in Reading, 29*(4), 383–399. <https://doi.org/10.1111/j.1467-9817.2006.0302.x>
- Liao, X., Zhu, X., & Zhao, P. (2022). The mediating effects of reading amount and strategy use in the relationship between intrinsic reading motivation and comprehension: differences between Grade 4 and Grade 6 students. *Reading and Writing, 35*(5), 1091–1118. <https://doi.org/10.1007/s11145-021-10218-6>
- Linacre, J. M., & Wright, B. D. (2001). *Winsteps (Version 3.61.1) [Computer software]*. Mesa Press
- Linacre, J. M. (2002a). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85–106

- Linacre, J. M. (2002b). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878
- Linacre, J. M. (2011). *A user's guide to WINSTEPS and MINISTEP: Rasch-model computer programs. Program manual 3.72.0*. Winsteps
- Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 4(2), 73–79. <https://doi.org/10.1027/1614-2241.4.2.73>
- Mason, L. H. (2013). Teaching students who struggle with learning to think before, while, and after reading: Effects of self-regulated strategy development instruction. *Reading & Writing Quarterly*, 29, 124–144. <https://doi.org/10.1080/10573569.2013.758561>
- Meyers, J., Lytle, S., Palladino, D., Devenpeck, G., & Green, M. (1990). Think-aloud protocol analysis: An investigation of reading comprehension strategies in fourth-and fifth-grade students. *Journal of Psychoeducational Assessment*, 8(2), 112–127. <https://doi.org/10.1177/07342829900800201>
- Miguel, J. P., Silva, J. T., & Prieto, G. (2013). Career Decision Self-Efficacy Scale — Short Form: A Rasch analysis of the Portuguese version. *Journal of Vocational Behavior*, 82(2), 116–123. <https://doi.org/10.1016/j.jvb.2012.12.001>
- Miyamoto, A., Pfost, M., & Artelt, C. (2019). The relationship between intrinsic motivation and reading comprehension: Mediating effects of reading amount and metacognitive knowledge of strategy use. *Scientific Studies of Reading*, 23(6), 445–460. <https://doi.org/10.1080/10888438.2019.1602836>
- Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, 94(2), 249–259. <https://doi.org/10.1037//0022-0663.94.2.249>
- Muijselaar, M. M. L., & de Jong, P. F. (2015). The effects of updating ability and knowledge of reading strategies on reading comprehension. *Learning and Individual Differences*, 43, 111–117. <https://doi.org/10.1016/j.lindif.2015.08.011>
- Muijselaar, M. M. L., Swart, N. M., Steenbeek-Planting, E. G., Droop, M., Verhoeven, L., & de Jong, P. F. (2017). Developmental relations between reading comprehension and reading strategies. *Scientific Studies of Reading*, 21(3), 194–209. <https://doi.org/10.1080/10888438.2017.1278763>
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill
- Parkinson, M. M., & Dinsmore, D. L. (2018). Multiple aspects of high school students' strategic processing on reading outcomes: The role of quantity, quality, and conjunctive strategy use. *British Journal of Educational Psychology*, 88, 42–62. <https://doi.org/10.1111/bjep.12176>
- Pereira-Laird, J., & Deane, F. (1997). Development and validation of a self-report measure of reading strategy use. *Reading Psychology*, 18(3), 185–235. <https://doi.org/10.1080/0270271970180301>
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1–15. [https://doi.org/10.1016/S0001-6918\(99\)00050-5](https://doi.org/10.1016/S0001-6918(99)00050-5)
- Prieto, G., & Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15(1), 94–100
- Reutzell, D. R., Smith, J. A., & Fawson, P. C. (2005). An evaluation of two approaches for teaching reading comprehension strategies in the primary years using science information texts. *Early Childhood Research Quarterly*, 20(3), 276–305. <https://doi.org/10.1016/j.ecresq.2005.07.002>
- Ribeiro, I., Dias, O., Oliveira, I. M., Miranda, P., Ferreira, G., Saraiva, M., Paulo, R., & Cadime, I. (2015). Adaptação e validação da escala Reading Strategy Use para a população portuguesa [Adaptation and validation of the Reading Strategy Use Scale for the portuguese population]. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 40(2), 25–36
- Rodrigues, B., Cadime, I., Viana, F. L., & Ribeiro, I. (2020). Developing and validating tests of reading and listening comprehension for fifth and sixth grade students in Portugal. *Frontiers in Psychology*, 11, 610876. <https://doi.org/10.3389/fpsyg.2020.610876>
- Rodrigues, B., Ribeiro, I., & Cadime, I. (2022). Reading, linguistic, and metacognitive skills: Are the relations reciprocal past the first school years? *Reading and Writing*. <https://doi.org/10.1007/s11145-022-10333-y>
- Rouquette, A., Hardouin, J. B., Vanhaesebrouck, A., Sébille, V., & Coste, J. (2019). Differential Item Functioning (DIF) in composite health measurement scale: Recommendations for characterizing DIF with meaningful consequences within the Rasch model framework. *Plos One*, 14(4), e0215073. <https://doi.org/10.1371/journal.pone.0215073>
- Santos, S., Cadime, I., Viana, F. L., Prieto, G., Chaves-Sousa, S., Spinillo, A. G., & Ribeiro, I. (2016). An application of the Rasch model to reading comprehension measurement. *Psicologia: Reflexão e Crítica*, 29(38), 1–19. <https://doi.org/10.1186/s41155-016-0044-6>

- Sondergeld, T. A., & Johnson, C. C. (2014). Using Rasch measurement for the development and use of affective assessments in science education research. *Science Education*, 98(4), 581–613. <https://doi.org/10.1002/sce.21118>
- Spörer, N., Brunstein, J. C., & Kieschke, U. (2009). Improving students' reading comprehension skills: Effects of strategy instruction and reciprocal teaching. *Learning and Instruction*, 19(3), 272–286. <https://doi.org/10.1016/j.learninstruc.2008.05.003>
- Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149–158. <https://doi.org/10.7334/psicothema2018.33>
- Tremblay, K. A., Binder, K. S., Ardoin, S. P., & Tighe, E. L. (2021). Third graders' strategy use and accuracy on an expository text: an exploratory study using eye movements. *Journal of Research in Reading*, 44(4), 737–756. <https://doi.org/10.1111/1467-9817.12369>
- Van Ammel, K., & Van Keer, H. (2021). Skill or will? The respective contribution of motivational and behavioural characteristics to secondary school students' reading comprehension. *Journal of Research in Reading*, 44, 574–596. <https://doi.org/10.1111/1467-9817.12356>
- Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning*, 1, 3–14. <https://doi.org/10.1007/s11409-006-6893-0>
- Veenman, M. V. (2011). Alternative assessment of strategy use with self-report instruments: a discussion. *Metacognition and Learning*, 6(2), 205–211. <https://doi.org/10.1007/s11409-011-9080-x>
- Vigil-Colet, A., Navarro-González, D., & Morales-Vives, F. (2020). To reverse or to not reverse Likert-type items: That is the question. *Psicothema*, 32(1), 108–114. <https://doi.org/10.7334/psicothema2019.286>
- Walker, C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment*, 29(4), 364–376. <https://doi.org/10.1177/0734282911406666>
- Wang, Y. (2016). Reading strategy use and comprehension performance of more successful and less successful readers: A think-aloud study. *Educational Sciences: Theory & Practice*, 16(5), 1789–1813. <https://doi.org/10.12738/estp.2016.5.0116>
- Weems, G. H., & Onwuegbuzie, A. J. (2001). The impact of midpoint responses and reverse coding on survey data. *Measurement & Evaluation in Counseling & Development*, 34(3), 166–176. <https://doi.org/10.1080/07481756.2002.12069033>
- Welter, V. D. E., Becker, L. B., & Großschedl, J. (2022). Helping learners become their own teachers: The beneficial impact of trained concept-mapping-strategy use on metacognitive regulation in learning. *Education Sciences*, 12(5), 325. <https://doi.org/10.3390/educsci12050325>
- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 189–194. <https://doi.org/10.1007/s10862-005-9004-7>
- Zhao, N., Wardeska, J. G., McGuire, S. Y., & Cook, E. (2014). Metacognition: An effective tool to promote success in college science learning. *Journal of College Science Teaching*, 43(4), 48–54

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.