



OPEN ACCESS

EDITED BY

Alejandro Javier Wainseboim,
CONICET Mendoza, Argentina

REVIEWED BY

Heikki Juhani Lyytinen,
Niilo Mäki institute, Finland
Maria Elena Rodriguez Perez,
University of Guadalajara, Mexico
Katherine Susan Binder,
Mount Holyoke College, United States

*CORRESPONDENCE

Barbara Weissenbacher
✉ barbara.weissenbacher@uni-graz.at

†These authors have contributed equally to
this work and share first authorship

RECEIVED 07 November 2024

ACCEPTED 27 January 2025

PUBLISHED 13 February 2025

CITATION

Sedlmayr P and Weissenbacher B (2025)
Reading comprehension assessment for
student selection: advantages of text
availability in terms of validity.
Front. Educ. 10:1524561.
doi: 10.3389/feduc.2025.1524561

COPYRIGHT

© 2025 Sedlmayr and Weissenbacher. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Reading comprehension assessment for student selection: advantages of text availability in terms of validity

Paul Sedlmayr[†] and Barbara Weissenbacher^{*†}

Department of Differential Psychology, Institute of Psychology, University of Graz, Graz, Austria

Reading comprehension assessments are widely used in university entrance exams across various fields. These tests vary in text content, answer format, and whether the text remains available while the questions are being answered. We explored how the availability of the text impacts the psychometric quality of the test and what is the best format for assessing reading comprehension in university admissions. We developed a test with four educational texts and 88 questions. A group of 107 university students was tested on two texts under two conditions: with and without text availability, using identical true/false questions for both scenarios. Results show similar difficulties in both versions, but slightly higher item-total correlations and internal consistencies in the *text-unavailable* condition, although this varied between texts. The *text-available* condition showed better validity, with expected correlations with participants' verbal intelligence and language scores. Overall, allowing text access during the test appears to be advantageous for assessing reading comprehension in university admissions.

KEYWORDS

reading comprehension, student selection, psychometric quality, validity, text availability

1 Introduction

Reading comprehension (or 'text comprehension') is crucial for various tasks in educational settings (Ferrer et al., 2017). It is therefore only logical that reading comprehension tests have considerable impact for academic success (Clinton-Lisell et al., 2022). Reading comprehension tests are frequently used as (part of) selection instruments in adults applying for various studies or for postsecondary course placement and advising (Clinton-Lisell et al., 2022; Orellana et al., 2024). Although reading comprehension tests are often used in adults, most of the research focuses on (younger) students, for example, examining secondary school students' PISA test performance (Yousefpoori-Naeim et al., 2023). In this study, we aim to learn more about reading comprehension tests as they are used with adults in admission tests and examine their psychometric properties using different test formats.

Reading comprehension entails linking ideas within a text and between the text and one's prior knowledge meaningfully, as described by the construction-integration model (Kintsch, 1998). Kintsch's (1998) model outlines three levels of text representation: the surface structure, which includes the text's exact words and grammar; the textbase, where these words form propositions and are interconnected; and the situation model, which merges the textbase with relevant background knowledge and personal experiences (Kendeou et al., 2016). The model includes both top-down (knowledge-driven) and bottom-up (word-based) processes in reading comprehension (Cain et al., 2017).

A number of competencies are related to reading comprehension. As we know from studies examining children as they learn to read, fundamental skills like phonological and morphological awareness and reading fluency are highly correlated with reading comprehension (Parrila et al., 2004; Quirk and Beem, 2012; Liu et al., 2024). Working memory, in terms of verbal processing and verbal storing, also plays an important role in reading comprehension (Daneman and Merikle, 1996). Besides, intelligence, specifically verbal reasoning, is associated with reading comprehension (Carver, 1990; Berninger et al., 2006). Meta-analyses examining the link between executive functions and reading comprehension find moderate correlations for working memory ($r = 0.29$; Peng et al., 2018) as well as executive functions in general ($r = 0.36$; Follmer, 2018).

Grammar and spelling, as other components of language proficiency, are also correlated with reading comprehension. Retelsdorf and Köller (2014) analyzed longitudinal data of German middle school students and found reciprocal effects between reading comprehension and spelling with a greater effect from reading comprehension to spelling compared to the opposite direction. In a meta-analysis, Zheng et al. (2023) found a high correlation between grammatical knowledge and reading comprehension. Besides, reading habits affect reading comprehension. Home literacy environment and children's reading comprehension are moderately correlated (Artelt et al., 2001; Dong et al., 2020). Throughout the lifespan, time spent reading and reading comprehension are positively correlated (Locher and Pfost, 2020). Recent research suggests that the benefit of frequent reading does not transfer from print to digital leisure reading habits (Altamura et al., 2023). As reading is a fundamental skill for educational purposes, it is logical that reading comprehension is related to academic achievement. Clinton-Lisell et al. (2022) meta-analytically examined the prognostic validity of reading comprehension tests on college performance and found a moderate relationship ($r = 0.29$).

Reading comprehension tests are typically structured to present test subjects with a written text, followed by questions based on the content. Tests are quite heterogeneous in terms of the nature of the text and response formats. Specifically, there are differences in text content and length, answer format (cloze versus question-and-answer), item format (open ended versus multiple-choice) and the availability of the text while answering questions (Cutting and Scarborough, 2006; Fletcher, 2006). The most frequently used item format is multiple choice. Although multiple-choice tests have often been criticized, they have been shown to have an advantage over open-ended questions in stimulating productive recall processes (Little et al., 2012).

Reading comprehension tests that employ question-and-answer formats can be designed in two ways: one where the target text is presented initially but disappears during the answering of text-based questions (text is not available), or another where the text remains accessible while the questions are being answered (text is available). As Ozuru et al. (2007) found out, text availability (versus non-availability) affects what aspect of reading comprehension is being measured. If the text remains available, participants can use a variety of strategies to engage with the text. They can switch back and forth from the text to the task or selectively search the text for information relevant to answering the questions (Reed et al., 2019). Even though reading the items before the text is a common strategy, previous studies have found no evidence supporting its effectiveness (Reed et al., 2019; Karsli et al., 2019). In contrast, if the text becomes unavailable during answering the questions, participants are incentivized to read the text more

carefully, potentially multiple times to memorize relevant information and ensure comprehension (Ferrer et al., 2017). This approach is very different to reading in everyday life (Schroeder, 2011).

In this study, we are interested in how text availability (versus non-availability) affects psychometric properties, and what format is more appropriate for assessing reading comprehension in university admissions tests. By presenting the same texts in two conditions (*text available* versus *text unavailable*), we examined item and scale properties as well as construct and criterion validity to answer the following research questions (RQ) and hypotheses (H):

- RQ1: How do item difficulties, item-total correlations and internal consistencies vary across conditions?
- RQ2: How are the two different test formats (text available versus text unavailable) intercorrelated?
- How does construct validity vary across conditions?
 - o H3a: In both conditions, grammar/spelling is expected to correlate with reading comprehension (Retelsdorf and Köller, 2014; Zheng et al., 2023).
 - o H3b: In both conditions, verbal reasoning is expected to correlate with reading comprehension (Berninger et al., 2006; Carver, 1990).
- How does criterion validity vary across conditions?
 - o H4a: In both conditions, academic achievement is expected to correlate with reading comprehension (Clinton-Lisell et al., 2022).
 - o H4b: In both conditions, reading habits are expected to correlate with reading comprehension (Locher and Pfost, 2020).

2 Method

2.1 Participants

Hundred and thirteen university students participated in the online study. Six participants were excluded from the statistical analysis, resulting in a sample size of $N = 107$. The excluded participants' test scores were at least 2.5 median absolute deviations below the sample median and within the range of guessing probability. The mean age of the final sample was 21.21 ($SD = 2.92$). The analysis was performed both with and without outliers. While the exclusion of outliers influenced the selection of specific items, it did not alter the overall pattern of effects reported in this study. Most students were in their first semesters ($Mdn = 2$), ensuring good comparability to admission test samples. This early semester distribution supports the prognostic validity of school grades as an external criterion. The great majority of participants were psychology students, who had the option of being reimbursed with course credits. Participants gave their informed consent prior to participating and the study procedure had been approved by the ethics committee of the University of Graz.

2.2 Instruments

2.2.1 Reading comprehension

We developed a reading comprehension test containing 4 texts on education-related topics and 88 true/false questions (items). Official

reports, for example government project reports (Federal Ministry of Education, Science and Research, 2024), served as sources for information. ChatGPT 4 (OpenAI, 2023) was used to assist in text creation.

In the sample tested, it cannot be assumed that there was any specific prior knowledge of the topics presented in the texts; for example, none of the topics covered is the subject of courses included in the curriculum of psychology. However, it cannot be ruled out that individuals may have prior knowledge due to their private interest in the topic, for example because the PISA results are regularly presented in the media.

The final texts were 500–600 words long, as suggested by Mashkovskaya (2013). The time limit for reading each text ranged from 4 to 5 min to account for slight differences in text length. This is 120 words per minute, which is twice as much time as the average reader needs (Brysaert, 2019). The purpose of this extended time is to allow slower readers to read the text in a meaningful way, and to allow participants to read sections repeatedly to memorize the content. Similarly, the time limit for answering items ranged from 5 to 6 min to account for differences in the number of items. When text and items were displayed in combination in the *text-available condition*, the time limits for reading and answering items were added up.

After each text, participants were presented with 20–24 questions on the text. For each question item, participants could choose between “Right,” “Wrong” or “Not in the text.” For each text, the answer categories were approximately balanced. The questions were based on the text and varied in the level of inference necessary to be answered correctly. An example of text and elements is shown in Figure 1. In the *text-unavailable condition*, participants first read the text and then answered the items without having the text at hand. In the *text-available condition*, the text and items were presented at the same time.

2.2.2 Verbal intelligence

For measuring verbal intelligence, we used the subscale “commonalities” of the German “Intelligence Structure Analysis” (ISA; Blum et al., 1998). The time-restricted subscale comprised 20 items; one item was excluded due to its negative item-total correlation. The internal consistency of the scale was low ($\alpha = 0.57$).

2.2.3 Grammar and spelling

For measuring German grammar and spelling skills, we used 25 self-developed single choice items assessing grammar and orthography ($\alpha = 0.70$). For example, participants were asked to choose the correct spelling of a difficult word from four different spellings, to choose the correct spelling in a gap sentence (from a list of 2–4 words, e.g., separate/combined spelling of words), or to choose the correct sentence with the correct tense and punctuation from four different sentences. As demonstrated in a pilot study, criterion validity could be assumed by moderate correlations with reading habits ($r = 0.29$, $p < 0.01$) and high school grades in German ($r = 0.51$, $p < 0.01$).

2.2.4 Reading habits

We asked the participants about their (average) weekly time spent on reading (open answer format). In addition, we asked on a five-point-scale (0 = never, 5 = very often) to rate how often they read (1) newspapers, (2) journals, (3) non-fiction books, (4) fiction books, and (5) online information. We averaged the scores for newspapers, journals and non-fiction books to obtain a mean score for “non-fiction” (compared to fiction and online information).

2.2.5 Previous academic achievement

Participants were asked to report their final (high school) grades in German and English. We computed participants’ average language grades by using arithmetical means.

Recent findings from the Programme for International Student Assessment (PISA) and the Health Behaviour in School-aged Children (HBSC) studies [...] show that most 15- to 16-year-old students in Austria feel comfortable in their schools. [...] However, there is a strong age effect in the relationship between pupils and their teachers: the relationship deteriorates with increasing age or length of time at school. Among 11-year-olds, 21% of female pupils and 15% of male students feel that their relationship with their teachers is very poor; among 17-year-olds, the figures are 46% and 55% respectively.

	Right	Wrong	Not in the text
A quarter of 17-year-old female students rate their relationship with teachers as very poor.		x	
The reason why students feel out of place is the poor relationship between teachers and students.			x

FIGURE 1
Text and item example (translated from German).

2.3 Procedure

After filling out demographic data, participants were presented with four texts. Each text was presented one at a time on the screen, followed by questions about the text. The entire text was presented on one page. Depending on the size of the screen, subjects had to scroll while reading. The font size was 14 pt. There was an option to zoom in or out through the browser. The first two texts were presented in the *text-available* condition. The third and fourth text were each presented in the *text-unavailable* condition. The reason for testing the *text-available* condition first was to prevent participants from misunderstanding the instructions (which included the information on whether the text would be available or not) and prematurely skipping the text in the *text-unavailable* condition. The four texts were randomly assigned to one of the conditions. The order within the condition was randomized as well. Next, validity information was collected by assessing (1) reading habits, (2) grammar and spelling and (3) verbal intelligence.

2.4 Data analysis

2.4.1 Item selection

First, items with an item-difficulty above 0.90 and below 0.10 were excluded. Second, we excluded items with a negative item-total correlation. Then, we excluded items with an item-total correlation below 0.15 in an iterative process. We have chosen this less strict limit because our analyses showed that items with an item-total correlation ≥ 0.15 increase the internal consistency of the scale.

Of 88 initial items, 45 remained in the *text-available* condition and 44 remained in the *text-unavailable* condition. Even though the number of items was similar in the two conditions, different sets of items emerged. Due to our randomization process, not every participant was presented with the same text in the same condition. Therefore, the items were selected specifically to each text and condition. For the item-total correlation, we used the correlation of an item with the scale mean excluding the item, as provided in the R psych package (Revelle, 2007).

2.4.2 Bootstrapping correlations

To examine the relationship between participants' test performance with external criteria, we used bootstrapping to estimate confidence intervals for Pearson's correlations. This approach was necessary due to a skewed distribution of participants' scores in both conditions. We used the percentile method of the R boot package (Canty and Ripley, 2024).

3 Results

3.1 Item characteristics

Across the four texts, the average item difficulty was $p_i = 0.70$ in both conditions. The average item-total-correlation was $r_{it} = 0.31$ in the *text-available* condition and $r_{it} = 0.36$ in the *text-unavailable* condition. Table 1 shows the item parameters (means, standard deviations and item-total-correlations) across the four texts (RQ1). Due to space limitations, we do not report values for each item; these can be retrieved from the [Supplementary material](#).

3.2 Scale characteristics

The average scale means were 0.75 (*text available*) and 0.72 (*text unavailable*); SD's: *text available*: $SD = 0.17$, *text unavailable*: $SD = 0.16$. The distribution was negatively skewed in both conditions, showing more extreme values in the *text-unavailable* condition. The kurtosis was higher in the *text-unavailable* condition than in the *text-available* condition. Due to the research design which meant that not every person received every text in each of the two conditions, it was not possible to compute the Cronbach's alpha coefficient across all four texts. Instead, the average internal consistency was computed. For the *text-available* condition, Cronbach's α was 0.65, for the *text-unavailable* condition, it was 0.72 (RQ1). Table 2 shows the scale characteristics for all four texts.

3.3 Validity

There was a moderate intercorrelation of $r = 0.34$ [0.10, 0.54] between the *text-available* condition and the *text-unavailable* condition (RQ2). Before item selection, the correlation between the conditions was moderate to high ($r = 0.51$ [0.33, 0.66]). To examine construct and criterion validity, we examined correlations with grammar and spelling, verbal intelligence, school grades and reading habits (see Table 3). In the *text-available* condition, we observed, as expected, significant correlations with verbal intelligence (hypothesis 3b) and language grades (hypothesis 4a). Unexpectedly, knowledge of grammar and spelling was not correlated to text comprehension (hypothesis 3a). Reading habits were not significantly correlated (hypothesis 4b). In the *text-unavailable condition*, none of the validity variables was significantly correlated with text comprehension (hypotheses 3a, 3b, 4a, 4b).

4 Discussion

In this study, we aimed to examine the psychometric properties of reading comprehension tests as they are used with adults in admission tests. More specifically, we were interested in how text availability (versus non-availability) affected psychometric properties including reliability and validity.

4.1 Item and scale characteristics

Comparing item and scale characteristics across conditions, we found that item difficulties - and therefore scale means - were comparably high between the two conditions. This result is consistent

TABLE 1 Item parameters (averaged over the four texts).

	Text available	Text unavailable
p_i	0.24–0.88 ($M = 0.70$)	0.36–0.88 ($M = 0.70$)
SD	0.33–0.51 ($M = 0.43$)	0.33–0.51 ($M = 0.43$)
r_{it}	0.17–0.54 ($M = 0.31$)	0.16–0.63 ($M = 0.36$)

p_i , item difficulty (corresponds to the item mean); r_{it} , correlation of an item with the scale mean excluding the item.

TABLE 2 Scale characteristics for each text and average.

	<i>N</i>	Number of items	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	α
Text 1 available	59	9	0.33	1.00	0.77	0.20	-0.73	-0.25	0.60
Text 1 unavailable	41	17	0.12	1.00	0.73	0.20	-1.05	0.70	0.80
Text 2 available	59	11	0.18	1.00	0.62	0.22	-0.16	-0.64	0.62
Text 2 unavailable	45	10	0.20	1.00	0.61	0.22	0.05	-1.12	0.65
Text 3 available	42	15	0.13	1.00	0.65	0.20	-0.54	0.04	0.72
Text 3 unavailable	57	9	0.00	1.00	0.70	0.22	-0.75	0.47	0.61
Text 4 available	54	10	0.00	1.00	0.81	0.20	-1.84	4.67	0.66
Text 4 unavailable	49	11	0.00	1.00	0.78	0.26	-1.43	1.45	0.80
Overall available	107	45	0.29	1.00	0.75	0.17	-0.75	-0.19	0.65
Overall unavailable	101	47	0.19	1.00	0.72	0.16	-1.08	1.00	0.72

It was not possible to calculate Cronbach's alpha for the overall scales because not every person received every text in each of the conditions. Therefore, an average score across the four texts was calculated for each of the conditions.

TABLE 3 Correlations with validity measures.

	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	$r_{\text{text available}}$	$r_{\text{text unavailable}}$
Grammar and spelling	8	24	17.06	3.73	0.15 [-0.03, 0.35]	0.18 [-0.04, 0.39]
Verbal intelligence	6	19	12.51	2.75	0.24 [0.04, 0.42]	0.11 [-0.10, 0.31]
Language grades	1	4	1.81	0.69	-0.26 [-0.44, -0.05]	-0.03 [-0.27, 0.19]
Total reading hours	0	25	8.23	6.89	-0.05 [-0.31, 0.21]	-0.10 [-0.36, 0.15]
Reading non-fiction	1	4	2.37	0.77	0.04 [-0.15, 0.24]	0.03 [-0.25, 0.06]
Reading fiction	1	5	3.25	1.26	0.12 [-0.06, 0.30]	-0.06 [-0.25, 0.11]
Reading online	2	5	4.25	0.79	0.17 [-0.04, 0.34]	0.06 [-0.15, 0.27]

CI's are based on bootstrapping with 1,000 samples (percentile method). Significant correlations shown in bold. Lower grades represent better achievement (1 = very good).

with Schroeder (2011) who found only marginal differences between the two conditions (in the sense that participants scored slightly higher when the text remained available). The item-total correlations were slightly higher in the *text-unavailable* condition which led to slightly higher internal consistencies. Due to the research design of our study which meant that not every person received every text in each of the two conditions, it was not possible to compute the Cronbach's alpha coefficient across all four texts. Looking at the internal consistencies of the individual texts, which are between 0.60 and 0.80, we can assume that a satisfactory internal consistency (exceeding the threshold of 0.70) would be achieved across all texts in both conditions; given that Cronbach's alpha increases with the number of intercorrelating items (Cortina, 1993). For practical use (e.g., student selection), two or more texts should be presented to increase the reliability.¹

4.2 Validity

Reading comprehension scores on the tests in the two conditions were moderately intercorrelated. The fact that the correlation was only

moderate indicates that both conditions do not measure (exactly) the same construct (i.e., the same competencies). This is consistent with previous findings, such as Ozuru et al. (2007), who concluded that text availability is related to different aspects of reading comprehension. Another explanation for the low correlation lies in the items and texts. The item-selection process resulted in distinct item sets for the two conditions. This is reflected in a decrease in the correlation from a moderate to high to just a moderate correlation. Furthermore, the same texts cannot be used to test each participant's performance in the two conditions. This induces additional variance, especially given potential interactions with participants' prior knowledge or interest in the topic of the text.

Examining construct and criterion validity, we found significant correlations with participants' verbal intelligence and language school grades in the *text-available* condition. Knowledge of grammar and spelling as well as reading habits were not correlated with reading comprehension in either of the conditions. No significant correlations were found with any of the validity measures in the *text-unavailable* condition.

Based on previous findings, we expected verbal reasoning—as a component of intelligence—to be associated with reading comprehension (Carver, 1990; Berninger et al., 2006). This expectation could only be met in the *text-available* condition. Unexpectedly, grammatical knowledge and spelling ability, which have been shown to be related to reading comprehension (Retelsdorf and Köller, 2014), were not significantly correlated in any of the conditions. An explanation for this contradictory result could be the small sample size

¹ Based on the results of this study, a reading comprehension test consisting of three texts and 44 items (total) in the *text-available* condition was created. The test showed acceptable reliability ($\alpha = 0.76$) in candidates applying for teacher education ($N = 2,839$).

of our study. To detect a small effect of $r = 0.15$, we would have needed a sample size of $N = 346$ (assuming a power of 0.80)².

Analyzing criterion validity, we found reading comprehension to be weakly to moderately related to previous academic achievement in the *text-available* condition. This result is in line with the meta-analysis of Clinton-Lisell et al. (2022) which indicated a moderate relationship between reading comprehension and college performance. Surprisingly, reading habits did not correlate significantly with reading comprehension. This was not only true for the weekly time spent on readings, but also for single reading activities, such as reading fiction, non-fiction and online information. Looking more closely, we see that reading fiction and online information showed almost significant correlations in the *text-available condition*. Therefore, it would be interesting to explore these relationships in a larger sample. Another reason for the lack of correlations could be the nature of the reading activities. Previous studies have shown that the benefit of frequent reading does not transfer from print to digital leisure reading habits (Altamura et al., 2023). We did not ask whether the time participants spent on reading activities was in print or digital form. Except for online information, all other reading activities measured could be consumed either offline or online. Therefore, it could be assumed that some of the participants' reports referred to print texts, while others referred to digital texts.

The finding that there were no correlations with criterion variables in the *text-unavailable* condition was surprising, especially with regard to verbal ability, since previous studies, such as the one by Schaffner and Schiefele (2013), found even higher correlations with intelligence in the *text-unavailable* condition than in the *text-available* condition (albeit in a paper-pencil study with middle school students in a between-subjects design). Especially with regard to possible sequence effects (see 4.3 Limitations), these effects should be re-examined in a replication study.

5 Limitations and conclusion

Since the first two texts were presented in the *text-available* condition and the third and fourth text were each presented in the *text-unavailable* condition, condition was confounded with order. The order presented here could have led to the following (undesired) effects: (1) If the subjects have already read several texts and answered questions, practice and carry-over effects could occur; this would mean that the subjects would achieve higher average scores on the last items, i.e., in the *text-unavailable* condition, compared to the items in the *text-available* condition at the beginning of the test. (2) If concentration and motivation decrease during the (online) test, this may lead to more errors and worse scores on average. However, we do not know which of these hypothesized effects occurred and to what extent and therefore cannot say how the sequence affected the results. These assumptions should be empirically tested in future studies. To avoid position effects, it would have been useful to randomly assign which text would be used in which experimental condition

and counterbalance its presentation. Alternatively, the two conditions could also be examined in a between-subjects design, which of course has its own disadvantages, such as the larger sample needed and confounding person variables (Charness et al., 2012).

Another problem is the restriction of variance in reading comprehension: since the tested sample consists of people who have already been admitted to university, it can be assumed that they have a higher level of reading comprehension than people who are just applying for admission to university, even if reading comprehension was not part of the selection process at the beginning of the course in the present sample.

Due to the small sample size, it was only possible to perform item analyses according to the classical test theory (Gulliksen, 1950). Further studies should include item response theory or structural equation modeling to test for differential item functioning or measurement invariance regarding, among others, demographic variables. Further studies could also include additional validity instruments; for example, it would be interesting to examine prior domain knowledge which is expected to correlate higher with reading comprehension if the text is unavailable (Ozuru et al., 2007) or self-regulation and decision-making which are supposed to influence performance on reading comprehension tests (Gil et al., 2015).

Since item and scale characteristics are comparable across the conditions, but validities are better in the *text-available* condition, the latter seems to be more suitable for measuring reading comprehension in university admission tests. In particular, the expected correlation with academic performance suggests that the higher validity in the *text-available* condition could be because the construct measured in this condition is more similar to academic requirements than in the *text-unavailable* condition. Given this more ecologically valid process of reading in the *text-available* condition, this approach provides a more reasonable choice for measuring reading comprehension.

Data availability statement

The raw data supporting the conclusions of this article are available under the following link: <https://osf.io/f953t/>.

Ethics statement

The studies involving humans were approved by Ethics Committee of the University of Graz. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

PS: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, Investigation. BW: Conceptualization, Data curation, Formal analysis, Methodology,

² In the aforementioned sample of teacher education candidates, reading comprehension was moderately correlated ($r = 0.41$) with spelling and grammar (using a parallel version of the test used in the present study).

Software, Validation, Visualization, Writing – original draft, Writing – review & editing, Project administration, Supervision.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. We would like to thank the University of Graz for funding the publication of this article.

Acknowledgments

The authors acknowledge the financial support by the University of Graz and would like to thank Maximilian Fromherz for his help in recruiting participants and collecting data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Altamura, L., Vargas, C., and Salmerón, L. (2023). Do new forms of Reading pay off? A Meta-analysis on the relationship between leisure digital Reading habits and text comprehension. *Rev. Educ. Res.* 95, 53–88. doi: 10.3102/00346543231216463
- Artelt, C., Schiefele, U., and Schneider, W. (2001). Predictors of reading literacy. *Eur. J. Psychol. Educ.* 16, 363–383. doi: 10.1007/BF03173188
- Berninger, V. W., Abbott, R. D., Vermeulen, K., and Fulton, C. M. (2006). Paths to Reading comprehension in at-risk second-grade readers. *J. Learn. Disabil.* 39, 334–351. doi: 10.1177/00222194060390040701
- Blum, F., Didi, H. J., Fay, E., Maichle, U., Trost, G., Wahlen, J. H., et al. (1998). Intelligenz Struktur Analyse (ISA). Ein Test zur Messung der Intelligenz. Frankfurt: Swets & Zeitlinger B. V., Swets Test Services.
- Brybaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *J. Mem. Lang.* 109:104047. doi: 10.1016/j.jml.2019.104047
- Cain, K., Compton, D. L., and Parrila, R. K. (2017). Theories of reading development. Amsterdam; Philadelphia: John Benjamins Publishing Company.
- Canty, A., and Ripley, B. D. (2024). Boot: bootstrap R (S-plus) functions.
- Carver, R. P. (1990). Intelligence and reading ability in grades 2–12. *Intelligence* 14, 449–455. doi: 10.1016/S0160-2896(05)80014-5
- Charness, G., Gneezy, U., and Kuhn, M. A. (2012). Experimental methods: between-subject and within-subject design. *J. Econ. Behav. Organ.* 81, 1–8. doi: 10.1016/j.jebo.2011.08.009
- Clinton-Lisell, V., Taylor, T., Carlson, S. E., Davison, M. L., and Seipel, B. (2022). Performance on Reading comprehension assessments and college achievement: a Meta-analysis. *J. Coll. Read. Learn.* 52, 191–211. doi: 10.1080/10790195.2022.2062626
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *J. Appl. Psychol.* 78, 98–104. doi: 10.1037/0021-9010.78.1.98
- Cutting, L. E., and Scarborough, H. S. (2006). Prediction of Reading comprehension: relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Sci. Stud. Read.* 10, 277–299. doi: 10.1207/s1532799xssr1003_5
- Daneman, M., and Merikle, P. M. (1996). Working memory and language comprehension: a meta-analysis. *Psychon. Bull. Rev.* 3, 422–433. doi: 10.3758/BF03214546
- Dong, Y., Wu, S. X.-Y., Dong, W.-Y., and Tang, Y. (2020). The effects of home literacy environment on Children's Reading comprehension development: a Meta-analysis. *Educ. Sci. Theory Pract.* 20, 63–82. doi: 10.12738/jestp.2020.2.005
- Federal Ministry of Education, Science and Research (2024). Available at: <https://pubshop.bmbwf.gv.at/> (Accessed January 27, 2024).
- Ferrer, A., Vidal-Abarca, E., Serrano, M.-Á., and Gilabert, R. (2017). Impact of text availability and question format on reading comprehension processes. *Contemp. Educ. Psychol.* 51, 404–415. doi: 10.1016/j.cedpsych.2017.10.002
- Fletcher, J. M. (2006). Measuring Reading comprehension. *Sci. Stud. Read.* 10, 323–330. doi: 10.1207/s1532799xssr1003_7
- Follmer, D. J. (2018). Executive function and Reading comprehension: a Meta-analytic review. *Educ. Psychol.* 53, 42–60. doi: 10.1080/00461520.2017.1309295
- Gil, L., Martinez, T., and Vidal-Abarca, E. (2015). Online assessment of strategic reading literacy skills. *Comput. Educ.* 82, 50–59. doi: 10.1016/j.compedu.2014.10.026
- Gulliksen, H. (1950). Theory of mental tests. Hoboken: John Wiley & Sons Inc.
- Karsli, M. B., Demirel, T., and Kurşun, E. (2019). Examination of different Reading strategies with eye tracking measures in paragraph questions. Hacettepe University journal of education. Available at: <https://api.semanticscholar.org/CorpusID:150720720> (Accessed January 20, 2025).
- Kendeou, P., McMaster, K. L., and Christ, T. J. (2016). Reading comprehension: Core components and processes. *Policy Insights Behav. Brain Sci.* 3, 62–69. doi: 10.1177/2372732215624707
- Kintsch, W. (1998). Comprehension: A paradigm for cognition. Cambridge, New York, NY, USA: Cambridge University Press.
- Little, J. L., Bjork, E. L., Bjork, R. A., and Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: fostering test-induced learning and avoiding test-induced forgetting. *Psychol. Sci.* 23, 1337–1344. doi: 10.1177/0956797612443370
- Liu, Y., Groen, M. A., and Cain, K. (2024). The association between morphological awareness and reading comprehension in children: a systematic review and meta-analysis. *Educ. Res. Rev.* 42:100571. doi: 10.1016/j.edurev.2023.100571
- Locher, F., and Pfost, M. (2020). The relation between time spent reading and reading comprehension throughout the life course. *J. Res. Read.* 43, 57–77. doi: 10.1111/1467-9817.12289
- Mashkovskaya, A. (2013). Der C-Test als Lesetest bei Muttersprachlern. Duisburg, Essen. Available at: <https://duepublico.uni-due.de/servlets/DocumentServlet?id=32859> (Accessed December 9, 2023).
- OpenAI (2023). ChatGPT. Available at: <https://chat.openai.com/chat>. (Accessed January 10, 2024).
- Orellana, P., Silva, M., and Iglesias, V. (2024). Students' reading comprehension level and reading demands in teacher education programs: the elephant in the room? *Front. Psychol.* 15:1324055. doi: 10.3389/fpsyg.2024.1324055
- Ozuru, Y., Best, R., Bell, C., Witherspoon, A., and McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cogn. Instr.* 25, 399–438. doi: 10.1080/07370000701632371

Generative AI statement

The authors declare that Gen AI was used in the creation of this manuscript. As mentioned in the Methods section, we used ChatGPT 4 (OpenAI, 2023) to assist in the text generation of the reading comprehension test.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2025.1524561/full#supplementary-material>

- Parrila, R., Kirby, J. R., and McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: longitudinal predictors of early Reading development? *Sci. Stud. Read.* 8, 3–26. doi: 10.1207/s1532799xssr0801_2
- Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., et al. (2018). A meta-analysis on the relation between reading and working memory. *Psychol. Bull.* 144, 48–76. doi: 10.1037/bul0000124
- Quirk, M., and Beem, S. (2012). Examining the relations between reading fluency and reading comprehension for english language learners. *Psychol. Sch.* 49, 539–553. doi: 10.1002/pits.21616
- Reed, D. K., Stevenson, N., and LeBeau, B. C. (2019). Reading comprehension assessment: the effects of Reading the items aloud before or after Reading the passage. *Elem. Sch. J.* 120, 300–318. doi: 10.1086/705784
- Retelsdorf, J., and Köller, O. (2014). Reciprocal effects between reading comprehension and spelling. *Learn. Individ. Differ.* 30, 77–83. doi: 10.1016/j.lindif.2013.11.007
- Revelle, W. (2007). Psych: Procedures for psychological, psychometric, and personality research. 2.4.6.26
- Schaffner, E., and Schiefele, U. (2013). The prediction of reading comprehension by cognitive and motivational factors: does text accessibility during comprehension testing make a difference? *Learn. Individ. Differ.* 26, 42–54. doi: 10.1016/j.lindif.2013.04.003
- Schroeder, S. (2011). What readers have and do: effects of students' verbal ability and reading time components on comprehension with and without text availability. *J. Educ. Psychol.* 103, 877–896. doi: 10.1037/a0023731
- Yousefpoori-Naeim, M., Bulut, O., and Tan, B. (2023). Predicting reading comprehension performance based on student characteristics and item properties. *Stud. Educ. Eval.* 79:101309. doi: 10.1016/j.stueduc.2023.101309
- Zheng, H., Miao, X., Dong, Y., and Yuan, D.-C. (2023). The relationship between grammatical knowledge and reading comprehension: a meta-analysis. *Front. Psychol.* 14:1098568. doi: 10.3389/fpsyg.2023.1098568