

Reading assessment in Brazil between the years 2014-2020: instruments and skills*

Margarete Gonçalves Macedo de Carvalho¹

ORCID: 0000-0002-0638-3343

Ana Cláudia de Souza²

ORCID: 0000-0002-0833-6903

Abstract

This paper presents the results of a systematic review on reading assessment. The problem that inspired the research was: What parameters, criteria, and assessment conditions can be distinguished in investigations on reading assessment in Brazil? Therefore, the goal was to identify, characterize, and discuss studies that were carried out on reading assessment in Brazil between 2014 and 2020. The focus of the study was on three central aspects of the 151 selected publications, according to the strict inclusion criteria described in the method: 1) the investigated cognitive processes, 2) the measured reading skills, and 3) the testing instruments used. As a criterion for analysis, we observed the research profiles (type, year, location, institution, graduate program, authorship, database) and their contents (focus, objective, target audience, reading evaluation instrument, measured skills, results). The analysis indicated that experimental studies, aimed mainly at students in the first grades of elementary school, were the most frequent, with decoding and its related skills as the most commonly investigated cognitive processes (word recognition, lexical access, and fluency). The results also showed that, in general, authors report the research methods with little detailing, which limits the identification of arguments to guide the decisions concerning elaboration or choice of instruments used. In addition, the lack of rigor or clarity of these aspects also makes it difficult for readers to understand the development of the studies and evaluate their reliability.

Keywords

Systematic review – Reading – Reading research – Reading assessment – Reading tests – Assessment instruments.

* The authors take full responsibility for the translation of the text, including titles of books/articles and the quotations originally published in Portuguese.

1- Instituto Federal de Santa Catarina (IFSC), Florianópolis, Santa Catarina, SC – Brasil. Contato: margarete.ifsc@gmail.com

2- Universidade Federal de Santa Catarina (UFSC), Florianópolis, Santa Catarina, SC – Brasil. Contato: ana.claudia.souza@ufsc.br



<https://doi.org/10.1590/S1678-4634202349259865eng>
This content is licensed under a Creative Commons attribution-type BY 4.0.

Introduction

Reading is a transversal topic that has been studied and debated by researchers from the most diverse areas of knowledge, among which Linguistics, mainly in two of its subareas, psycholinguistics and applied linguistics, as well as education, psychology, speech-language and hearing sciences, medicine, and others. We can justify its transversality because, as a multidimensional phenomenon, reading cannot be studied in only one area since it involves biological and cultural aspects that make it a complex activity. For that reason, it shall be investigated and understood from multiple perspectives and different delimitations.

This study assumes a psycholinguistics perspective, which observes reading considering the cognitive, metacognitive, and affective relations established between the individual and the stimulus to which s/he is exposed - the text - in a given environment and condition, pursuing a comprehension purpose. Given the nature of the writing system, reading is a cultural activity generally learned through systematized teaching and learning processes, usually in a school setting.

Synthetically and considering the lens used here to observe reading, it is possible to identify two large dimensions regarding its processing: linguistic decoding and comprehension, that is, the graphemic-phonemic mapping, which leads to word recognition, and meaning production, achieved by the relationship between the various textual layers and readers' relevant prior knowledge (HOOVER; TUNMER, 1992; VIANA, 2009; PERFETTI; LANDI; OAKHILL, 2013). Considering the complexity of readers' encounter with the text, we understand reading as a competence consisting of a set of skills, as explained by Souza, Seimetz-Rodrigues, and Weirich (2019, p. 166, our translation):

To say that reading is a competence implies [...] considering it as a set of non-compulsory and non-spontaneous skills (or, at least, not spontaneous until readers learn it) that need to be pondered upon, developed, and practiced to know what to do and how to do when reading, so that there is a chance of reaching the purpose of some meaning production (such sort depends on what the reader, teacher, or instructor wants), in other words, that there is the possibility of triggering a mental representation and retextualization process of the writing.

The skills required for reading range from the most basic to the most complex levels are activated depending on the reading or the reading comprehension goals and the specific procedural, declarative, and conditional knowledge readers have. The objectives can be related to identification, analysis, elaboration, discussion, and synthesis aspects, among others. The required skills involve word identification (through decoding or not), morphological-syntactic computation, basic meanings access, meaning production beyond the text line, and different orders of inferential processes that affect both the word unit and larger textual units and their interrelations.

As societies have become more dependent on writing³, new levels of reading competence have been required of individuals, which has generated the need to elaborate mechanisms to evaluate them. Not only do schools use reading assessment as a means^{of}

3- This use of writing means the secondary modality of language, its notational representation.

diagnosis, a starting point for teaching procedures, and students' ranking, but also other areas of society use it for selection, remediation, and research on its impact on humanity.

Historically, the assessment of reading skills dates back to the seventh century in China, where it was used, along with other intelligence tests, in the examination system for civil service and later for admission to universities. The same occurred in Britain and the U.S. in the twentieth century (GARDNER, 1996). In the school setting, according to Pearson and Hamm (2005), it dates back to the early days of school; however, as a formal activity, we only started to discuss it in the 20th century, just before World War II. Initially, we sought to assess oral reading; later, silent reading, moving on to standardized tests and reading speed evaluation - encouraged by behaviorism - until the early 1970s.

Reading assessment has developed mainly from new technologies for investigating the components of reading comprehension proposed by Davis in 1944 (PEARSON; HAMM, 2005), going through several other assessment techniques focused on decoding and comprehension.

In the Brazilian context, when undertaking a bibliographic review on research that shows the trajectory of reading assessment, a gap is noted due to the scarcity of literature in the area, which consists of sparse scientific articles. The identified books deal with socio-historical information and discursive and political aspects of reading from the nineteenth century onwards, emphasizing literacy, readers' subjective aspects, or the book history (MORTATTI *et al.*, 2014; LAJOLO; ZILBERMAN, 2019) without addressing the assessment issue. The theoretical approaches on reading, published from 1970 on, coincide with the American ones and are based on them; however, in Brazil, in recent decades, there has been an emphasis on social-historical and interactionist aspects of reading, to the detriment of cognitive ones, especially in the school setting.

The studies point to some aspects that we must observe when assessing reading. First, we must take into account the readers we intend to evaluate. Also, we must consider what the assessment and reading purposes are, the type of knowledge (declarative, procedural, conditional), and the linguistic level (phonological, lexical, morphological, syntactic, semantic - both phrasal and interphrasal -, textual, discursive or pragmatic) to be assessed. In reading comprehension, we also must ponder upon the aspect we intend to consider, the level of complexity we want to grasp, and, finally, what resources we have available to carry out the assessment.

It is worth noting that reading is an activity that is always indirectly assessed due to the psychological processes involved in it and the fact that no instrument is complete enough to evaluate it (SOUZA; CARVALHO, 2019), especially concerning comprehension. It is necessary, therefore, to have a set of instruments and techniques with different formats, items, texts, and supports, among others, that comprise the entire extent of the construct, from basic to higher skills, as long as the purpose of the assessment, the assessed people, and the conditions under which it will occur are considered (ALDERSON, 2000).

Based on the above considerations, this research aims to identify, characterize, and discuss research conducted in Brazil on reading assessment between the years 2014 and 2020, focusing on cognitive processes investigated, the reading skills measured, and the testing instruments used to outline parameters, criteria, and conditions that researchers

have considered relevant for the assessment. In sum, through a systematic review study, we aim to investigate the following question: What parameters, criteria, and conditions for assessment can be distinguished in reading assessment investigations in Brazil?

Method

The systematic review method aims to know the scientific production on a given topic, allowing researchers to situate themselves as to the state of the question (NÓBREGA-THERRIEN; THERRIEN, 2004).

To carry out this review, we conducted the first search using the following keyword combinations, using Boolean operators: assessment, reading assessment, assessment instrument, reading test, instrument, tes?, test, testing, AND reading OR reading comprehension, OR reading performance⁴, according to the syntax allowed by the search sources of the databases. The databases of the Virtual Health Library (VHL - Lilacs and Pepsic) and the Scientific Electronic Library On-line (SciELO), the Portal de Periódicos Capes, and the Brazilian Digital Library of Dissertations and Theses (BDTD) were checked in June and July 2020 and revised in January and April 2021, as they are considered relevant databases for national publications in psychology, interfaces, and related areas, such as psycholinguistics, education, psychopedagogy, and speech-language and hearing sciences, which are especially interested in the topic.

Once we identified the studies, we applied the following inclusion criteria: studies that assessed typical reading, published between 2014 and 2020, whose target participants were Brazilian Portuguese speakers from 6 years old on or in the 1st grade of elementary school, and whose reading tests or instruments used Brazilian Portuguese. In the case of articles derived from dissertations or theses already identified and selected, we included only those whose scope was not the same as the original work. For the pre-selection, we analyzed the presence of any descriptors or correlates in the title and keywords, and after that, we referred to the objectives announced in the abstracts. We excluded publications with one or more characteristics: being present in more than one database, presenting discussion focus unrelated to the descriptors, not being empirical research, or deriving and discussing the same aspects of a dissertation or thesis already selected.

We restricted the search to articles, theses, and dissertations published in the last seven years before the writing of this article, considering a previous publication by Dias *et al.* (2016), which covered the years 2009 to 2013. Therefore, the cut-off period from 2014 to 2020 is justified because there is a previous systematic review that covers the topic until 2013, and it is not feasible to include research from the year in which the search was revised for confirmation purposes: January to April 2021. Thus, initially, we identify 8,482 publications from the descriptors, 7,615 of them in the *Portal de Periódicos Capes*, 439 in the VHL (Lilacs), 92 in the VHL (Pepsic), 46 in the BDTD, and 290 in SciELO. We based the last step of the selection on the content of the publications, achieving a total of 151 articles, theses, or dissertations, which were under detailed analysis, and the readers of this paper can identify them by the letter T (*Trabalho*, which means publication), followed

4- Please, refer to the original article in Portuguese to check the keywords because English does not allow for certain variations.

by the order number - for instance, T1, T2, and so forth. Readers can access the analyzed publications, whose titles we ordered alphabetically, in: https://drive.google.com/file/d/1tnNejzi73mPKscxvCw5x9i1nBDrLmotS/view?usp=drive_link.

After the selection process, data were tabulated, discriminating and analyzing the following variables: profile - 1) Type of research, 2) Year, 3) Journal (in the case of articles), 4) University, 5) Graduate Program (GP), 6) Authorship: authors (for articles), author and supervisor (for dissertations and theses), 7) Database; and content - 1) Focus of interest, 2) Objective, 3) Target audience (participants), 4) Instrument used to assess reading, 5) Reading skills measured, 6) Results. Finally, based on what was reached by the analysis of the variables - whose information we gathered to help understand the target data and the research as a whole - the sample was examined, according to categories of testing instruments and reading skills, observing the parameters used in the construction or selection of the instrument. A critical discussion of the material follows the analysis.

It is noteworthy that, as clarified in the presentation of the problem and the goals of this bibliographic research carried out by the systematic review method, this is a mapping study of the state of the question, not a propositional one about the delineation of parameters, criteria, and conditions relevant to the assessment of reading comprehension. Carvalho (2022) developed such a propositional study.

Results

Research profile

As for the textual genre, of the 151 studies in the final sample, 108 were articles, 29 were master's theses, and 14 were doctoral dissertations. The graduate programs (GP) undertook 43 pieces of research on the subject, with the Letters⁵ and Cognitive Psychology programs producing the most, with eight and six papers, respectively. Regarding GPs, the institutions that published the most were PUCRS, UFPE, and USP (six, six, and five, in this order). The Linguistics programs were responsible for only 9.5% of the research, divided into the subfields of Psycholinguistics and Applied Linguistics; of these, the university with the highest representation was UFPE. Of the 43 dissertations and theses, BDTD hosted 42 and *Portal de Periódicos da Capes one*.

Regarding authorship, most of the professors who researched the topic supervised one work each, except for Alina G. Spinillo (UFPE), who supervised five, Augusto Buchweitz (PUCRS) and Simone Aparecida Capellini (UNESP), three each, and Camila Domeniconi (UFSCAR), Denise P. Cardoso (UFS), Janáina Weissheimer (UFRN), Maria Regina Maluf (PUCSP) and Vera W. Pereira (PUCRS), who supervised two papers each. The researchers who participated in more publications in journals in the period are from psychology: Acácia Aparecida A. dos Santos (18), Neide de B. Cunha (eight), Katya Luciane de Oliveira (seven), Adriana S. Ferraz (four), Patrícia S. Lucio (four), Márcia Maria P. E. da Mota (four), and from speech-language and hearing sciences: Simone Aparecida Capellini (seven),

5- Only those whose GPs' names were Letters.

Maria Silvia Carnio (five), and Aparecido José C. Soares (four). The journals that published the most papers were those in psychology (62) and speech-language and hearing sciences (27). The other areas were letters/linguistics (eight), neuropsychology/health (seven), psychopedagogy (four), and mathematics (one). The results show the interdisciplinary nature of reading research and its evaluation, with psychology and its interfaces specially interested in the topic.

Table 1 shows the total amount of studies distributed per year:

Table 1– Piece of research per year of publication

Year	2014	2015	2016	2017	2018	2019	2020	Total
N.	25	29	28	26	16	19	08	151

Source: From our elaboration (2021).

The average, per year, from 2014 to 2017 was 27 works, and the interest in investigating aspects involving the evaluation of reading suffered a decrease in 2018, with a drop of around 60%. Regarding the research developed within the GPs, only six defenses occurred. In 2020, there was an even more evident drop, possibly due to the restrictions imposed by the Covid-19 pandemic, an event that somehow affected all activities in most countries in the world. In that year, we could not locate any dissertations or theses on the subject, and the research reported through the publications was probably carried out in previous years.

Research content

In general, we found that the focus of interest that remained the most over the years was the validation of instruments and the effects of interventions. The majority of the studies (132) established their target audience based on schooling, and the stage of education with the highest number of studies was elementary education (121), especially in the early grades. Four studies investigated secondary school students; one investigated youth and adult education (EJA⁶) students, and eleven investigated higher education (HE) students. Those that selected participants by age group (17) did so according to the following distribution: 6 years (one), 6-12 years (11), 9-14 (three), 14-16 years (one), and adults and elderly (six). In addition, one study aimed to investigate elementary Portuguese language teachers. The numbers show an emphasis on research related to initial reading, with a sharp decrease concerning high school, EJA, and HE students, and even individuals in their specialty areas or those who are not students anymore.

We outlined the study methods based on the environment, the theoretical approach, and the data collection technique (GIL, 2017). Thus, we counted 137 experimental studies, four

6- EJA is the acronym in Brazilian Portuguese for *Educação de Jovens e Adultos*.

action studies, two descriptive/exploratory studies, two case studies, and six mixed-method studies (involving non-participant observation, action research, and experimental research).

The instruments used for testing were quite varied: researchers used around 64 different types to measure reading skills, and some used more than one. In table 2, we present the four main instruments or testing techniques used in the selected studies.

Table 2 – Most frequent assessment instruments in the studies

Instruments	N.	Skills assessed in the studies	References
1) Various reading comprehension tests - RCT (elaborative answer, multiple-choice, written and oral tests)	38	Literal and inferential reading comprehension, information location, and retrieval	The authors elaborated the tests.
2) Cloze test	37	Reading comprehension	TAYLOR (1953) (original)
3) <i>Teste de Desempenho Escolar</i> – Reading subtest (TDE)*#	18	Isolated word recognition	STEIN (1994)
4) <i>Provas de Avaliação dos Processos de Leitura</i> (PROLEC)#	11	Decoding, lexical decision, fluency (speed and accuracy), and initial reading comprehension	CAPELLINI; OLIVEIRA; CUETOS (2010)

* This test consists of three subtests: arithmetic, writing, and reading. For this work, only the reading subtest was of interest.

Tests originally produced in Portuguese had their original names kept in this article.

Source: From our elaboration (2021).

Of the instruments listed in table 2, 75 assess only reading comprehension and 18, word recognition; the remainder evaluate two or more processes.

The studies targeting students in the early grades of elementary education investigated skills related to decoding, fluency (accuracy, speed, and prosody), literal and inferential interpretation, the relationship between phonological, morphological, syntactic, and metatextual awareness, and reading comprehension, and vocabulary, using mainly the TDE - Reading Subtest, and PROLEC. For students in the final grades of elementary school, the emphasis was on the ability to draw inferences. Studies whose participants were high school or HE students emphasized aspects related to reading comprehension.

Table 2 shows that the majority of those studies that investigated reading comprehension used elaborative answer and multiple-choice reading comprehension tests (38) - henceforth RCT - and cloze tests (37) almost exclusively (only 11 and five, respectively, used them in conjunction with another comprehension assessment instrument). Five RCTs used the same test model, utilizing more than one text or task. In most, researchers developed these two types of tests specifically for the research in question.

The primacy of RCT and cloze tests in the research analyzed contrasts with the results of Dias *et al.* (2016), who found that TDE was ahead of the cloze test as the most used. We can explain this by the authors' selection of papers: limited to articles and included atypical development populations.

In the studies in which the investigated skills corresponded to reading comprehension, we sought to verify the impact on it of factors of quite varied orders: linguistic (connectives, prosody, syntactic simplification, textual type plus genre, vocabulary, and fluency), individual (gender, age, socioeconomic status), cognitive (memory, metacognition, auditory processing), sociocultural (*e.g.*, reading and literacy practices), pedagogical (teaching method, interventions, order of tasks) and affective (motivation).

Regarding objectives, the studies that sought to validate instruments under development assessed psychometric characteristics related to internal structure, construct, content, and criterion validity, as translating and adapting research instruments. Some have combined the search for evidence of validity with participants' reading performance observation. Those dedicated to verifying the impact of interventions on the development or improvement of reading skills exhibited a diversity of objectives: to investigate, verify, and analyze the effects of phonological awareness activities and the teaching of grapheme-phoneme mapping; working memory in the increase of reading skills; syllabic coloring technique, verbal and non-verbal resources of the comic book genre; synonyms teaching; reading comprehension remediation program; teaching methods. Others aimed to characterize, describe, and observe the evolution, compare and correlate reading performance, verify reading comprehension level, identify decoding and reading comprehension difficulties, investigate the relationship between extralinguistic factors and reading, and others.

Description and analysis

We analyzed the instruments used and the reading skills measured in the selected works considering that the purpose of a psychometric reading assessment instrument is to measure, albeit indirectly, the individual's mastery of the processes required in reading through the manipulation by the researcher of tasks, text factors, and reading situation or condition. With this in mind, we considered the following categories for analysis: a) cognitive processes - approaches, strategies, or cognitive purposes used by readers in the course of their involvement with the text (context or intertext) aiming at the accomplishment of the proposed task; b) required skill - triggering of procedural knowledge; c) difficulty level - reading stage accessed through the test; d) test input characteristics - stimulus, item type, and response format; e) explanation for choosing the test - criteria that determined the selection or creation of the testing instrument; f) information on validation - types of search for evidence of validity to which the researcher submitted the instrument.

Reading comprehension tests

The 38⁷ RCTs used to assess reading comprehension were varied. Researchers based some on the *Prova Brasil* (three), *Indicador de Alfabetismo Funcional - INAF* (two),

7- T1, T2, T5, T7, T8, T9, T11, T16, T26, T27, T31, T32, T38, T39, T43, T44, T48, T49, T50, T53, T54, T60, T69, T70, T71, T72, T79, T80, T107, T113, T118, T123, T125, T126, T129, T131, T135, and T148.

Exame Nacional de Desempenho do Estudantes - ENADE (one), Avaliação Nacional da Alfabetização - ANA (one), Provinha Brasil (one), and tests developed by other authors, such as Mahon (2002) and Oakhill and colleagues (2005) (two). In other cases, they developed specific tests for the research.

The publications analyzed generalized the cognitive process investigated as reading comprehension, and in most cases, they defined only reading skills. We observed inference-making skills in 26 studies; literal interpretation in 10; information localization in four; metacognitive strategies in five; and information evaluation/reflection in four.⁸ Regarding the level of difficulty, four studies reported that the RCT was appropriate to educational stage and grade, two only controlled the level of difficulty of the texts, one study highlighted that specialists evaluate the degree of texts' difficulty, questions, and alternatives of response, and, contrary to what we expected from scientific research, 31 publications did not indicate whether the researchers adjusted the test to the reading stage to be accessed.

The input characteristics varied, as shown in table 3:

Table 3 – RCT input characteristics

Stimulus	Support	Text format	Item format	Total
visual	printed	continuous/combined	Multiple choice	07
visual	printed	continuous/combined	Elaborative answer	08
visual	printed	continuous/combined	Multiple choice and elaborative answer	07
visual	printed	continuous	Oral elaborative answer	04
visual	printed	continuous	Oral questions and answers	04
visual and auditory	printed	continuous	Oral questions and answers	04
visual and auditory	printed	continuous/combined	Oral elaborative answer	02
visual	virtual	continuous/conceptual mapping	Not mentioned	01
		Not mentioned		01
				38

Source: From our elaboration (2021).

It is remarkable that of the 38 studies that used RCT, five did so digitally; some used questions asked and answered orally during the text reading, with programmed interruptions. Those investigations that used visual and auditory stimuli justified the researcher's oral reading while participants would read the text to facilitate the process for students who may have decoding difficulties. We inferred the support in four papers since they did not mention it.

Regarding the justification for the choice of the instrument, 27 studies did not indicate it, a fact that is surprising since one of the criteria for choosing a particular tool is the analysis of its efficiency for the objective proposed for the evaluation. The absence

8- In this research, we consider the skills of localization and evaluation/reflection as cognitive processes, according to the Programme for International Student Assessment (Pisa).

of justification may reveal a lack of clarity about the characteristics and qualities that led to its choice. Those studies that have done so justified that the RCT is a resource that helps the reading comprehension performance classification. One explained that it allows for a more controlled assessment of various comprehension processes. Some studies specifically clarified the choice of the elaborative answer item format, stating that they favor argumentation, the relationships between text and social reading practices, and the identification of processing that requires more cognitive effort on the part of the readers. The authors of one publication justified the multiple-choice format, explaining that it allows investigating aspects related to the contextual meaning of words, identifying the author's objectives and point of view, literal and inferential comprehension of the text, and being objective and easy to correct. This justification is doubtful since we can observe the same skills through elaborative answers. Some based their choice on the reliability of a standardized test, such as *Prova Brasil*. In one study, the authors explained that, although they considered the questions test not ideal for checking comprehension, they chose the instrument to check whether readers could evoke literal and inferential aspects of the text and not only decode it.

Regarding evidence of test validity, only six of the 38 studies mentioned it, indicating content validity and the degrees of precision and discrimination of the items or just stating that they found this evidence, two of which dedicated themselves to the construction and validation of instruments.

Cloze test

Of the 37⁹ studies that used cloze tests, 20 used the following instruments developed by Santos (2005) and validated for elementary school: *A princesa e o fantasma* and *Uma vingança infeliz*, for the first grades, and *Coisas da natureza*, for the final ones. Four studies used the text *Desentendimento* by Veríssimo (1995); one, the instrument *O campinho* by Spinillo and Mahon (2007); there was the development of six texts specifically for the investigation, one of them part of a subtest of the *Instrumento de Avaliação da Leitura Inicial (IAL-I)* and another one was a cloze of sentences test.

As for the cognitive processes investigated, 18 studies did not detail them; however, we inferred that they would be those of information elaboration and interpretation since they evaluate the construction of the situational model of the text by the reader. Eight did not allude to the term cognitive process; of these, two used the expression comprehension process, one reading process, and five did not refer to the term as conceived in this review. In eight studies, reading comprehension was considered either a construct composed of various cognitive skills and processes or a skill to be measured by the instrument or specific task demands.

In 31 studies, the skill verified or the demand required by the task was reading comprehension, in the sense of identifying global understanding of the text, based on the correct gap-filling. Two studies verified inference-building skills; two, reading

⁹- T6, T15, T18, T20, T23, T24, T25, T33, T34, T35, T36, T37, T38, T39, T40, T41, T42, T45, T46, T47, T51, T66, T72, T75, T78, T87, T94, T92, T98, T102, T107, T114, T117, T122, T131, T133, and T151.

comprehension at the textual structure level, creating a demand for syntactic and lexical choice; one investigation analyzed immediate local processing; and another, interpretation.

Regarding the level of difficulty, that is, the reading stage accessed through the test, 15 indicated that the instruments took into account the level of schooling or school grade, and two just pointed out that the level of complexity is low, easy, or hard, without justifying it; the other publications made no mention. The stimuli and support characteristics were visual (written) and printed, respectively.

As for the test format, five used multiple-choice cloze, which is a variation that allows the reader to fill in the gap from several alternatives, according to Santos Burochovitch and Oliveira (2009). In one of the papers, it was unclear whether the response format was oral or whether the type of test was post-oral reading cloze. The other considered Taylor's (1953) traditional cloze.

Once more, contrary to the expectation from scientific research, of the 37 studies, 16 did not justify the choice of the instrument used. Those who did, emphasized that the option for the cloze test was due to its practical, fast, and reliable measure, flexible and easy to use, correction and calculation of scores, as well as its low cost and the fact that there is already clear evidence of its validity, especially criterion validity, relative to the school grades evaluated.

As for information on instrument evidence of validity, 17 publications made some mention, 16 made reference to evidence of criterion validity for the tests or texts, three of content, two studies of construct, three of convergent, five cited aspects related to internal consistency, and three to accuracy indexes, three studies mentioned the existence of evidence of instrument validity, without making it clear what it is, and three highlighted aspects of validity of the original cloze model.

Teste de Desempenho Escolar (TDE) - Reading subtest

Eighteen¹⁰ studies used the Teste de Desempenho Escolar (TDE) - Reading subtest to assess the proposed skills. In 14 publications, the cognitive process observed was decoding; the other four did not clarify. They measured isolated-word reading skills.

Five studies justified the difficulty level based on theoretical assumptions regarding the distribution of the items by grade level. Considering that it was a single instrument with a defined protocol, the stimuli were generally visual, the support printed, and the response format was oral.

Six studies did not present the justification for choosing the TDE. Among those that did, three emphasized that it is an instrument built for Brazilian schoolchildren. Most of them emphasized the test validity evidence aspect, especially criterion validity. One justified the interest in standardizing an edition that assesses students in the elementary school final grades, and another that the use occurred to define the sample composition. Half of the studies did not refer to the instrument validity evidence. Those that did, as mentioned, highlighted its criterion validity.

10- T30, T41, T42, T58, T64, T73, T76, T77, T82, T89, T95, T96, T97, T98, T102, T104, T116, and T132.

Provas de Avaliação dos Processos da Leitura - PROLEC

Of the 151 studies, eleven¹¹ used the PROLEC and PROLEC - SE-R as measurement instruments. The PROLEC battery tasks aim for students from 2nd to 5th grade and PROLEC - SE-R is for students from 6th to 9th grade. Cuetos, Rodrigues, and Ruano developed the original standardized instrument in 1996, and Capellini, Oliveira, and Cuetos adapted it to Brazilian Portuguese in 2014.

Five studies that used PROLEC/PROLEC - SE- R aimed to observe the decoding process, four, the reading comprehension process, two both, and one, to translate and adapt the instrument. Those focusing on decoding investigated the ability to read words and pseudowords (three), to identify letters and lexical processes (two), frequent and infrequent words (one), and syntactic and semantic processes (one). Those that focused on comprehension measured skills related to literal and inferential interpretation. As for the difficulty level, ten did not mention it, and one argued that it is of a low degree of difficulty and low variability of scores, which would show unsatisfactory internal validity in his results on the *Prova de Avaliação de Textos* (PROLEC-T).

Since it was a printed material, the stimuli and support did not vary. The items' format was elaborative answers; of these, three studies cited that the answer was oral, three did not inform whether it was oral or written, and five did not indicate the format type.

Three studies justified the choice of the instrument; one of them, based on its ability to assess the different processes and sub-processes involved in reading; another, based on its objective, explained that the choice was due to the need to seek evidence of validity and to evaluate the standardization of the Brazilian version published in the manual. The last one based itself on the wide acceptance of the instrument by education and health professionals. Concerning the evidence of the instrument validity, only two studies mentioned it, one specifying the criteria and the other the external validity.

Instruments used in more than two studies

In addition to those described above, among the 64 types of instruments used in the investigations, the tools used in more than two works were: morphological awareness assessment tasks^{12/13}, lexical decision tasks^{14/15}, *Protocolo de Avaliação da Compreensão de Leitura* (PROCOMLE)¹⁶, by Cunha and Capellini (2019), oral reading tests¹⁷, *Questionário de Avaliação da Consciência Metatextual* (QACM)¹⁸, by Santos and Cunha (2012), *Teste*

11- T10, T17, T56, T97, T99, T106, T108, T111, T112, T127, and T128.

12- T34, T35, T41, T51, T91, and T117.

13- Instruments built by authors such as Sá (1999, 2006), Paula (2007), Mota and Brilhante (2012), Nunes, Bryant, and Bindman (1997), adapted by Justi and Roazzi (2012), and Guimarães and Mota (2016).

14- T9, T14, T42, T63, T91, and T100.

15- Two used the semantic priming paradigm, one the syllabic priming technique, and one the morphological, orthographic, and unrelated priming.

16- T28, T84, T105, T120, and T132.

17- T15, T27, T48, T55, T80, and T141.

18- T20, T21, T66, T75, and T117.

*de Competência de Leitura Silenciosa de Palavras e Pseudopalavras (TCLPP)*¹⁹, by Seabra and Capovilla (2010), *Teste de Nomeação Automatizada Rápida (RAN)*^{20/21}, by Denckla and Rudel (1974), retelling²², word reading tests²³ - three based on Salles (2005), *Teste Contrastivo de Compreensão Auditiva e de Leitura*²⁴, by Capovilla and Seabra (2013)²⁵ and the *Teste de Leitura: Compreensão de Sentenças (TELCS)*²⁶, adapted from Lobrot (1967, 1980) by Vilhena and colleagues (2016).

These studies focused on the cognitive processes of decoding, lexical access, metalanguage, and reading comprehension. The skills measured were: word recognition, fluency, metatextual, morphological, and morphosyntactic awareness, and, in reading comprehension, literal and inferential interpretation. As for the task difficulty level, few studies have mentioned it. Most studies cited the input characteristics, although some did so with little detail.

Again, it is striking that very little research has justified the instrument choice. The same happened with its validity evidence. Those studies that explain it highlighted content, construct, and criterion validity, the latter, especially about the participants' school stage.

Based on the objectives proposed for this systematic review research and the problem it aims to answer, a discussion of the most relevant aspects follows.

General discussion

Most of the selected studies were scientific articles aiming to find evidence of validity for instruments they built or used to assess reading in Brazil or verify the effects of interventions on participants' reading performance. Given this, we found that the most used methodological approach was experimental, although in many cases presenting no explanation regarding the environment control level, as required by this research method. It is worth noting, however, that the non-verbalization of control does not necessarily indicate a lack of compliance with this requirement for experimental research, although experiments require clarity in methodological detailing.

The level of schooling defined the target participants of the research, and the priority in formal education was either the elementary school first grades (in most cases) or the age of the group, which also coincided with the educational stage. This trend in reading research focused on the school environment and the first years of schooling is also a feature of international databases. This tendency reflects that researchers have dedicated little attention to investigating the behavior of proficient adult readers, specialized reading, and, consequently, higher levels of reading competence, which has had repercussions on the

19- T15, T30, T65, T86, and T119.

20- T41, T93, T106, and T132.

21- Most commonly used acronym in English: *Rapid Automated Naming*. Considering the focus of this systematic review, we selected only those publications that employed the letters' subtest.

22- T60, T83, T94, and T113.

23- T5, T9, T60, and T130.

24- T59, T86, and T119.

25- One study employed TCLPP – II – elementary school 6th to 9th grades (without associated images).

26- T4, T42, T115, and T149.

limited number of investigations focused on strategic reading comprehension. According to PISA, these levels of reading competence range from skills that are basically about explicitly locating small pieces of information to those that enable broad and detailed comprehension and require the integration of one or more texts (OCDE, 2017).

Most studies, especially those on decoding and metalanguage, used commercial instruments. Considering that most of them investigated factors related to the decoding process, the skills measured were limited to words and pseudowords reading, fluency, phonological, morphological, and syntactic awareness, and, concerning comprehension, initial literal and inferential interpretation skills. Therefore, the studies highlighted the most elementary levels of processing related to reading. This choice by the researchers may reflect the lower-than-expected reading proficiency rates of Brazilian students in the last decade and the identification that the problems that motivate them are related to the initial stage of learning to read. (OCDE, 2018; BRASIL, 2020).

Studies that focused on reading comprehension constructed their respective data collection instruments. One aspect that draws attention is that half of those that investigated reading comprehension aspects used only one tool, which widely was the cloze test. Alderson (2000, p. 206) warns of the inadequacy of this type of assessment: "Good reading tests are likely to employ a number of different techniques, possibly even on the same text, but certainly across the range of texts tested" because distinct tools allow us to observe distinguishable reading skills (OAKHILL; CAIN; ELBRO, 2017). Moreover, according to Alderson (2000, p. 9), "Text constructors [...] must also consider the level of meaning that they believe readers ought to 'get out of' a particular test when assessing 'how well' they have understood the text in question."

Maintaining the discussion on cloze tests, when referring to what researchers observed of cognitive processes, the publications were limited to mentioning reading comprehension in general, without specifying which comprehension component. Studies have shown, however, that this test format relates mainly to decoding ability (NATION; SNOWLING, 1997). Keenan, Betjemann, and Olson (2008) illustrated this in a cloze item in which failure to decode only one word led participants to an incorrect answer. Alderson (2000, p. 7) explains that cloze techniques may induce "some readers to read in a particular way (paying close attention to individual words, for instance, or reading the text preceding the gap, but not the following text)." This procedure makes it difficult to generalize performance to a particular reading ability. Another issue raised by the author, when it comes to reading comprehension, is that cloze does not assess whether the reader has read a text critically or only passively, which limits the view of comprehension derived from the product evaluated by this type of instrument.

Most papers did not inform what evidence of validity they found to endorse the use of the instrument in the research or justify its choice. Based on the information contained therein, we inferred that the credibility of a given tool was in the fact that many studies had employed it. This finding corroborates the statement of Keenan, Betjemann, and Olson (2008, p. 294) concerning what is recognized theoretically about reading comprehension but which is unconfirmed in research:

Comprehension is a complex cognitive construct, consisting of multiple component skills. Even though this complexity is recognized theoretically, when it comes to assessment, there is a tendency to ignore it and treat tests as if they are all measuring the same “thing.” This is reflected in the fact that researchers who measure comprehension rarely give information on why they chose the particular test that they used. Implicit in this behavior is the suggestion that it does not really matter which test was used because they are all measuring the same construct.

Another aspect to consider is that, especially concerning the RCTs, we found no arguments in the research analyzed to justify the choice of the base texts or the format, content, and structure of the items based on the observed cognitive process and measured skill. When selecting a text for an RCT, it is necessary to remember that its complexity is one of the factors that most affect the interaction of each reader with the individual assessment questions. It is required to balance the items in such a way as to know what level of reasoning (for example, inferring, analyzing, synthesizing) will be necessary to answer them so that all the intended levels are contemplated (ACHIEVE.ORG, 2019).

At first, there was little detail or insight into the method of the studies analyzed, which limited the identification of parameters, criteria, or conditions defined to guide the decisions to create or choose the instruments used. In addition, we understand that the lack of rigor or clarity of these aspects also makes it difficult for the text reader to comprehend how the researchers developed the studies and their reliability. This fact may have occurred because most papers are articles whose journal publication has limited writing space. Although the lack of methodological information may not necessarily disqualify a piece of research, it is relevant to consider that the reader accesses just what the writer tells about the investigation, not accessing its development process. Hence the importance of research reports being as transparent as possible about their theoretical and methodological foundations.

Final remarks

The motivating question of the present investigation was: What parameters, criteria, and evaluation conditions can be distinguished in the studies on reading evaluation in Brazil? To answer it, the general objective of this bibliographic research was to map and discuss, according to the systematic review method, what researchers have investigated in Brazil on reading assessment, which instruments they have been using in the research, cognitive processes they have been focusing, reading skills they have been measuring and what the outcomes have been indicating, from their methods, as parameters for the evaluation of reading.

Probably, this research did not cover many works due to the parameters of the search engines of the databases, which consist, in general, of their titles, authors, or topics/keywords. In this sense, it did not include research on reading or reading comprehension that does not focus on testing, reading assessment, or reading comprehension, as it does not appear in the items cited. For future systematic reviews, we suggest reversing the

search to select publications emphasizing the investigation of reading skills or reading comprehension and only indirectly the instruments used.

Knowing what has been investigated, from what theories and methods, as well as results found about the object of interest, namely, the evaluation of reading, allowed us to gather information from other research to enable understanding the framework of scientific knowledge produced in recent years on the topic, as well as to guide the planning of new investigations, indicating possible research gaps.

Based on the results found in the investigations on reading assessment in Brazil between 2014 and 2020, we conclude that the reading level most strongly investigated is that of initial reading, the stage for which we observed the majority of available instruments, and it is not necessary to develop a data collection instrument specifically to investigate more advanced or in-depth reading comprehension processes at this educational stage. The most commonly used data collection instruments are reading comprehension tests (RCTs), cloze tests, the *Teste de Desempenho Escolar - Reading Subtest (TDE)*, and the *Provas de Avaliação dos Processos de Leitura (PROLEC)*.

A relevant aspect to highlight is the difficulty of finding clear information about the methodological decisions of a significant part of the research. There is a lack of details and information to justify and argue in favor of a particular decision. The aim of this analysis was not to evaluate the quality of the research but to provide evidence from the literature about which research has been investigating reading assessment in Brazil in the last decade and how researchers have been conducting their studies concerning methodological aspects. It was not within the scope of this study to propose parameters and criteria for reading assessment. Carvalho (2022) conducted a study with this purpose.

Finally, we emphasize the importance of systematic review studies in all areas of scientific investigation, considering it is a research method that seeks to collect empirical evidence based on clearly delineated parameters, which allow reproducibility and fit predetermined eligibility criteria to answer an investigative question. The study reported here presents a mapping of the research conducted in Brazil in the last decade, which has focused on reading assessment, highlighting methodological aspects and other relevant information to recognize what researchers have been investigating and how they conduct their studies. This type of information provides subsidies for future research on the same topic bringing light to other researchers by indicating the art state. It is a research method that favors the advancement of science through the synthesis of research already developed.

References

ACHIEVE.ORG. **A framework to evaluate cognitive complexity in reading assessments**. [S. l.: s. n.], 2019. Disponível em: <https://files.eric.ed.gov/fulltext/ED603577.pdf> Acesso em: 02 fev. 2021.

ALDERSON, J. Charles. **Assessing reading**. Cambridge: Cambridge University Press, 2000.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP. **Ideb – resultados 2019**. Brasília, DF: INEP, 2020. Disponível em: https://download.inep.gov.br/educacao_basica/portal_ideb/documentos/2020/Apresentacao_Coletiva_Imprensa_Saeb_2019.pdf. Acesso em: 12 fev. 2021.

CAPELLINI, Simone A.; OLIVEIRA, Adriana Marques de; CUETOS, Fernando. **Prolec: provas de avaliação dos processos de leitura**. São Paulo: Casa do Psicólogo, 2010.

CARVALHO, Margarete Gonçalves Macedo de. **Avaliação da compreensão leitora: parâmetros e critérios essenciais para a construção de instrumentos**. 2022. 222 f. Tese (Doutorado em Linguística) – Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2022. Disponível em: <https://tede.ufsc.br/teses/PLLG0877-T.pdf> Acesso em: 09 jan. 2023.

DIAS, Natália M. *et al.* Avaliação da leitura no Brasil: revisão da literatura no recorte 2009-2013. **Psicologia: Teoria & Prática**, São Paulo, v. 18, n. 1, p. 113-128, abr. 2016. Disponível em: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1516-36872016000100009&lng=pt&nrm=iso Acesso em: 01 jun. 2020.

GARDNER, Howard. **Inteligências múltiplas**. Porto Alegre: Artmed, 1996.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2017.

HOOVER, Wesley A. TUNMER William E. The components of reading. *In*: THOMPSON, Brian *et al.* **Reading acquisition processes**. Philadelphia: Multilingual Matters, 1992. p. 1-17.

KEENAN, Janice M.; BETJEMANN, Rebecca S.; OLSON, Richard K. Reading comprehension tests vary in the skills they assess: differential dependence on decoding and oral comprehension. **Scientific Studies of Reading**, London, v. 12, n. 3, p. 281-300, 2008. <https://doi.org/10.1080/10888430802132279>

LAJOLO, Marisa; ZILBERMAN, Regina. **A formação da leitura no Brasil** [recurso eletrônico]. 2. ed. São Paulo: Unesp Digital, 2019. Edição do Kindle.

MORTATTI, Maria do Rosário L. *et al.* (org.). **Sujeitos da história do ensino de leitura e escrita no Brasil**. São Paulo: SciELO: Unesp, 2014. Edição do Kindle.

NATION, Kate; SNOWLING, Margareth. Assessing reading difficulties: the validity and utility of current measures of reading skill. **British Journal of Educational Psychology**, London, v. 67, p. 359-370, 1997. Disponível em: <https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2044-8279.1997.tb01250.x> Acesso em: 16 out. 2020.

NÓBREGA-THERRIEN, Sílvia Maria; THERRIEN, Jacques. Os trabalhos científicos e o estado da questão: reflexões teórico-metodológicas. **Estudos em Avaliação Educacional**, São Paulo, v. 15, n. 30, jul./dez. 2004. Disponível em: <http://publicacoes.fcc.org.br/index.php/eae/article/view/2148/2105> Acesso em: 01 mar. 2019.

OAKHILL, Jane; CAIN, Kate; ELBRO, Carsten. **Compreensão de leitura: teoria e prática**. Trad. Adail Sobral. São Paulo: Hogrefe, 2017.

OCDE. Organização para a Cooperação e Desenvolvimento Econômico. **PISA 2015 assessment and analytical framework**: science, reading, mathematic, financial literacy and collaborative problem solving. Paris: PISA: OECD, 2017. <http://dx.doi.org/10.1787/9789264281820-en>

OCDE. Organização para a Cooperação e Desenvolvimento Econômico. **PISA 2018 results**. Paris: PISA: OECD, 2018. Disponível em: https://www.oecd.org/pisa/PISA-results_ENGLISH.png Acesso em: 12 fev. 2021.

PEARSON, P. David; HAMM, Diane N. The assessment of reading comprehension: a review of practices - past, present, and future. *In*: PARIS, Scott G.; STAHL, Steven A. **Children's reading comprehension and assessment**. Oxfordshire: Routledge, 2005. p. 13-70.

PERFETTI, Charles A.; LANDI, Nicole; OAKHILL, Jane. A aquisição da habilidade de compreensão da leitura. *In*: SNOWLING, Margaret J.; HULME, Charles. **A ciência da leitura**. Porto Alegre: Penso, 2013. p. 519-538. Edição do Kindle.

PRIMI, Ricardo; MUNIZ, Monalisa; NUNES, Carlos Henrique S. S. Definições contemporâneas de validade de testes psicológicos. *In*: HUTZ, Claudio S. (org.). **Avanços e polêmicas em avaliação psicológica**: em homenagem a Jurema Alcides Cunha. São Paulo: Casa do Psicólogo, 2009. p. 243-265.

SANTOS, Acácia Aparecida Angeli; BUROCHOVITCH, Evely; OLIVEIRA, Katya. L. (org.). **Cloze**: um instrumento de diagnóstico e intervenção. São Paulo: Casa do Psicólogo, 2009.

SOUZA, Ana Cláudia de; CARVALHO, Margarete Gonçalves Macedo. Professor-leitor: o que dizem as pesquisas brasileiras. **(Con)textos Linguísticos**, Vitória, v. 13, p. 156-175, 2019.

SOUZA, Ana Cláudia de; RODRIGUES, Cristiane Seimetz; WEIRICH, Helena Cristina. Ensinar a estudar ensinando a ler: potências dos roteiros de leitura. *In*: SOUZA, Ana Cláudia *et al.* (org.). **Diálogos linguísticos para a leitura e a escrita**. Florianópolis: Insular, 2019. p. 164-200. Disponível em: <https://insular.com.br/produto/dialogos-linguisticos-para-a-leitura-e-a-escrita-nao-comercializado/> Acesso em: 09 jan. 2023.

STEIN, Lillian M. **TDE - Teste de desempenho escolar: manual para aplicação e interpretação**. São Paulo: Casa do Psicólogo, 1994.

TAYLOR, Wilson L. Cloze procedure: a new tool for measuring readability. **Journalism Quarterly**, Columbia, v. 30, p. 415-433, 1953. Disponível em: <https://www.gwern.net/docs/psychology/writing/1953-taylor.pdf> Acesso em: 03 dez. 2021.

VIANA, Fernanda Leopoldina. **O ensino da leitura**: a avaliação. Lisboa: Direcção-geral de Inovação e de Desenvolvimento Curricular: Ministério da Educação, 2009. Disponível em: http://repositorium.sdum.uminho.pt/bitstream/1822/31558/1/ensino_leitura_avaliacao.pdf Acesso em: 03 dez. 2021.

*Received on January 7th, 2022
Reviewed on November 19th, 2022
Approved on February 6th, 2023*

Editor: Prof. Dr. Marcos Sidnei Pagotto Eusébio

Margarete Gonçalves Macedo de Carvalho holds a Ph.D. in linguistics from the Federal University of Santa Catarina. She is a federal public servant working with educational issues at the Instituto Federal de Santa Catarina.

Ana Cláudia de Souza holds a Ph.D. in linguistics from the Federal University of Santa Catarina. She is a full professor at the Universidade Federal de Santa Catarina in the Department of Metodologia de Ensino and the Graduate Program in Linguistics.