

Vera Lúcia Orlandi Cunha¹
Simone Aparecida Capellini¹

Construction and validation of an instrument to assess the reading comprehension of students from the third to the fifth grades of elementary school

Construção e validação de instrumento de avaliação da compreensão de leitura para escolares do terceiro ao quinto ano do Ensino Fundamental

Keywords

Reading
Comprehension
Evaluation
Educational assessment
Schooling

Descritores

Leitura
Compreensão
Avaliação
Avaliação educacional
Escolaridade

ABSTRACT

Purpose: In this study, we aimed at building and validating an instrument to assess reading comprehension, with the purpose of characterizing the reading profile and detecting comprehension difficulties among students from the third to the fifth grades of elementary school. **Methods:** Participants were 378 students, divided into three groups. Their comprehension of micro- and macrostructural literal and inferential propositions that composed two expository texts and two narrative texts were assessed by means of multiple-choice questions. **Results:** The data analyzed statistically yielded Cronbach's alpha values showing internal consistency in the four texts applied to the three groups. **Conclusion:** It was possible to verify that the students had fewer errors as the school years progressed and that each type of text posed a particular difficulty to the students.

RESUMO

Objetivo: Este estudo teve como objetivo construir e validar um instrumento de avaliação da compreensão de leitura a fim de caracterizar o perfil de leitura e detectar dificuldades de compreensão em escolares do terceiro ao quinto ano do Ensino Fundamental. **Métodos:** Participaram 378 escolares divididos em três grupos para avaliação da compreensão de proposições literais e inferenciais de micro e macroestruturas de dois textos expositivos e dois textos narrativos por meio de questões de múltipla escolha. **Resultados:** Os dados analisados estatisticamente indicaram valores do teste alfa de Cronbach apresentando consistência interna nos quatro textos aplicados para os três grupos. **Conclusão:** Foi possível constatar que os escolares apresentaram menor número de erros com o aumento da escolarização e que cada tipo de texto apresentou uma dificuldade específica para os escolares.

Correspondence address:

Vera Lúcia Orlandi Cunha
Rua das Camélias, 218, Jardim das Flores,
Palmital (SP), Brazil, CEP: 19970-000.
E-mail: verafono@uol.com.br

Received: 03/19/2013

Accepted: 19/11/2013

CoDAS 2014;26(1):28-37

Study carried out at the Universidade Estadual Paulista "Júlio de Mesquita Filho" – UNESP – Marília (SP), Brazil. (1) Philosophy and Sciences School at Universidade Estadual Paulista "Júlio de Mesquita Filho" – UNESP – Marília (SP), Brazil.

Financial support: Coordination for the Improvement of Higher Education Personnel (CAPES).

Conflict of interest: nothing to declare.

INTRODUCTION

The ability to read and comprehend texts is fundamental in our daily lives, and, when associated with education, it is invested with a greater importance for being a component intrinsic to the cognitive development of students⁽¹⁾.

Problems in reading comprehension are an obstacle in the learning process, as all school tasks (in Portuguese, Geography, History, and also Mathematics) require the students to read and extract important information that they need for learning. Students with comprehension problems are not able to perform these tasks and fall behind in comparison to their classmates⁽²⁾.

The comprehension process entails two types of processes – basic and high level. Basic processes distinguish themselves from high-level processes because the latter require better mental elaboration abilities. Some important abilities considered as basic level are work memory and lexical processes (for instance, knowledge of orthographic structures), whereas high-level abilities involve inferring (about information that is not explicit in the text or that encompasses knowledge on the topic acquired previously) and monitoring what is being understood⁽³⁾.

The cognitive representation involved in comprehending written texts has been classified into three levels: “surface code,” which preserves the exact vocabulary and the syntax of sentences; “source text,” which contains the propositions of the text displayed in a manner so as to preserve the meaning but not the vocabulary or the exact form of the text (it includes the necessary inferences to establish textual coherence); and “situational model,” which is the mental representation of what is explicitly mentioned or inferentially suggested in the text (the majority of inferences generated on the course of text comprehension are part of the situational model)^(4,5).

Thus, during reading, an individual performs an analysis and a comparison between information extracted from the text (both with regard to word decoding and recognition and to text comprehension) and previously stored information. In order to understand a text profoundly, the reader must formulate two types of inference – literal inferences, relating ideas within or between sentences, and implicit inferences, connecting ideas in order to complement information that is not explicit, thus incorporating previous knowledge and experiences. This process is necessary for the assemblage of a text’s model of mental representation^(4,6).

Texts are classified in many types or categories. However, the great majority of investigations is fundamentally centered on expository and narrative texts due to the fact, at least in part, that students are more exposed to these types of texts since childhood and during the educational process⁽⁵⁾.

Based on the aforementioned studies, the instrument proposed here has the purpose of objectively detecting in which level lies the student’s difficulties, that is, whether at the microstructural and macrostructural levels and/or in

relation to literal or inferential information contained in narrative and expository texts, by means of multiple-choice questions (see examples in Appendix 1).

We opted for multiple-choice questions because it is a means of evaluating the abilities involved in comprehension that has the advantage of enabling the investigation of aspects related to the contextual meaning of words and the identification of an author’s intention and point of view, while passing through literal and inferential information⁽⁷⁾.

The use of multiple-choice questions to assess student knowledge is becoming more frequent, and a large part of this popularity is due to the objectivity and the easiness to correct the answers given. However, this assessment method has some limitations insofar as attentiveness is necessary when elaborating enunciations and choices for each question⁽⁸⁾. Thus, to put together the instrument proposed in this study, we drew on norms of elaboration for psychometric instruments that utilize multiple-choice questions^(9,10).

An assessment instrument is basically founded on statistical values that indicate its precision (accurate values with regard to reliability and stability of the results) and validity (assurance that the test measures what it purports to measure)^(9,10). The validity of an instrument is defined by considering if it indeed measures what, supposedly, it must measure. The statistical analyses conducted on an instrument, in its entirety or by each individual item, are performed under the assumption that it is unidimensional. This implies that all items of an instrument measure the same construct. Moreover, the concept of accuracy comprises different aspects of a test, but all of them refer to the extent to which an individual’s scores remain identical on different occasions; for instance, the scores obtained at a first and at a second moment for the same individuals⁽⁹⁾.

Therefore, in order to confer validity to an assessment instrument, it needs to be constructed, from the outset, based on criteria and norms determined by ethics and by the quality of the instruments that will be later used by researchers and clinical professionals, who obtain data that need to rest on reliable comparative evidence.

Considering the aspects presented thus far, in this study, we aimed at constructing and validating an instrument to assess the reading comprehension of expository and narrative texts, with the purpose of characterizing the profile and detecting reading comprehension difficulties among students from the third to the fifth grades of elementary school.

METHODS

This study was conducted after being submitted and approved by the Ethics Committee of the Faculty of Philosophy and Sciences at Universidade Estadual Paulista “Júlio de Mesquita Filho” (FFC/UNESP), Marília (SP), under report number 1881/2008.

The instrument was constructed according to the following phases:

- Phase 1: from the beginning, the construction of this instrument was based on the theoretical foundations of the Model of Reading Comprehension Mental Representation^(4,5), presented in the introduction of this article.
- Phase 2: stages of the construction of the instrument to assess reading comprehension – survey on assessment instruments available in Brazil and abroad; selection of the population; selection of the texts by teachers and professionals who study reading comprehension; elaboration of the questions to assess textual comprehension; judgment of the questions by professionals who study reading comprehension; preliminary verification of the applicability of the instrument through application of the Brainstorming Technique⁽⁹⁾ and a pilot study; definition of the final contents of the instrument: two expository texts (E1 and E2) and two narrative texts (N1 and N2), each with four literal questions and four inferential questions (two concerning the microstructure and two related to the macrostructure).
- Phase 3: validation of the instrument. The analyses required to validate the instrument proposed were based on internal consistency, described through the analysis of the comparison of correct answers given by the students during two applications of the instrument (first collective application and second collective application).
- Phase 4: characterization of the profile of the students from the third to the fifth grades of elementary school regarding reading comprehension, by means of analyzing the incorrect answers given by these students in the first collective application.

Application procedures

We conducted two collective applications of the assessment instrument proposed in this study, with the purpose of characterizing the reading comprehension profile of the students, as well as comparing their performances in the first application and the second application in order to verify the consistency of the answers in both applications.

Participants of the first collective application

We adopted the following inclusion criteria: students whose parents or legal guardians had signed the informed

consent; students without sensory, motor or cognitive deficiencies enrolled in the school; students without decoding difficulties; students who participated in the application of the four texts that compose the assessment instrument.

The participants were 378 students enrolled in elementary school, who were divided in three groups: Group I (GI): 102 third-year students (of both sexes, within the 8-year age range); Group II (GII): 121 fourth-year students (of both sexes, within the 9-year age range); Group III (GIII): 155 fifth-year students (of both sexes, within the 10-year age range).

Participants of the second collective application

The participants of the second application were 138 students selected out of the 378 individuals who participated in the first application, allocated as follows: Group I (GI): 34 third-year students (of both sexes, within the 8-year age range); Group II (GII): 46 fourth-year students (of both sexes, within the 9-year age range); Group III (GIII): 58 fifth-year students (of both sexes, within the 10-year age range).

During both applications, all students were submitted to a reading assessment with the four texts that compose this instrument and their respective questions.

RESULTS

The reliability levels of the values observed were statistically analyzed through Cronbach's alpha coefficient, as displayed in Table 1, where the significance for each expository text (E1 and E2) and for each narrative text (N1 and N2) is indicated for groups GI (third year), GII (fourth year), and GIII (fifth year).

The results indicated internal consistency, with Cronbach's alpha coefficient levels between 0.600 and 0.700 for the four texts applied to the three groups of students, as per Table 1.

Since the results obtained were consistent in all groups, we analyzed the percentage of correct answers among the groups with regard to each text applied through a likelihood ratio test.

Graph 1 displays the percentage of correct answers for each group, concerning each text applied.

The data found in Graph 1 show that the percentage of correct answers effectively increased from GI to GII to

Table 1. Cronbach's alpha coefficient and significance values of the four texts applied with the three groups studied

Group	Aspect	n	α	p-value
I	E1 and E2	102	0.661	<0.001*
I	N1 and N2	102	0.648	<0.001*
II	E1 and E2	121	0.683	<0.001*
II	N1 and N2	121	0.640	<0.001*
III	E1 and E2	155	0.627	<0.001*
III	N1 and N2	155	0.657	<0.001*

*Statistical test used: Cronbach's alpha coefficient

Caption: E1, expository text 1; E2, expository text 2; N1, narrative text 1; N2, narrative text 2; α = Cronbach's alpha coefficient

GIII, that is, the performances improved from GI to GIII. This finding suggests that, as the students proceed from one school year to the next, they develop the cognitive and linguistic processes necessary for textual comprehension. This indicates, therefore, differences in the students'

performances in accordance with their schooling and better performances in more advanced years.

In order to verify the reproducibility of the instrument, that is, if the participants' answers remained the same in two applications of the instruments in similar situations, we used McNemar's test. With this test, it was possible to analyze if there was equivalence between both applications of the instrument (first and second collective applications) in relation to the variables of interest. This comparison was conducted individual/individual, with the purpose of verifying each participant's behavior during each application.

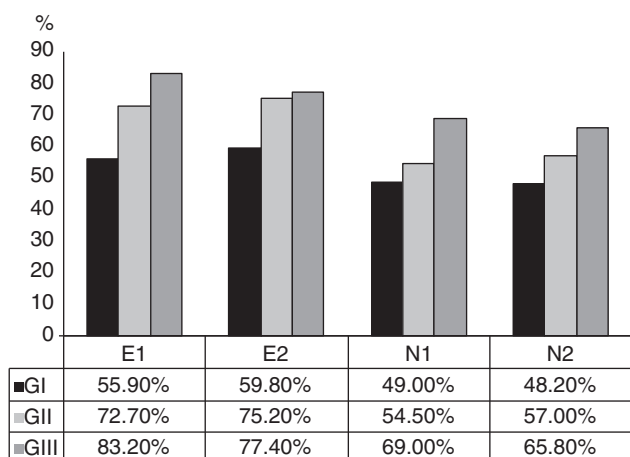
Graphs 2 to 5 present the significance level (p-value) found for each group with regard to each question upon comparison of the first and the second applications of the expository texts E1 and E2.

The data presented in Graphs 2 and 3 display statistically significant differences concerning the expository texts E1 and E2 in only two literal questions (one of microstructure and of one macrostructure).

Graphs 4 and 5 present the p value found for each group with regard to each question upon comparison of the first and the second applications of the narrative texts N1 and N2.

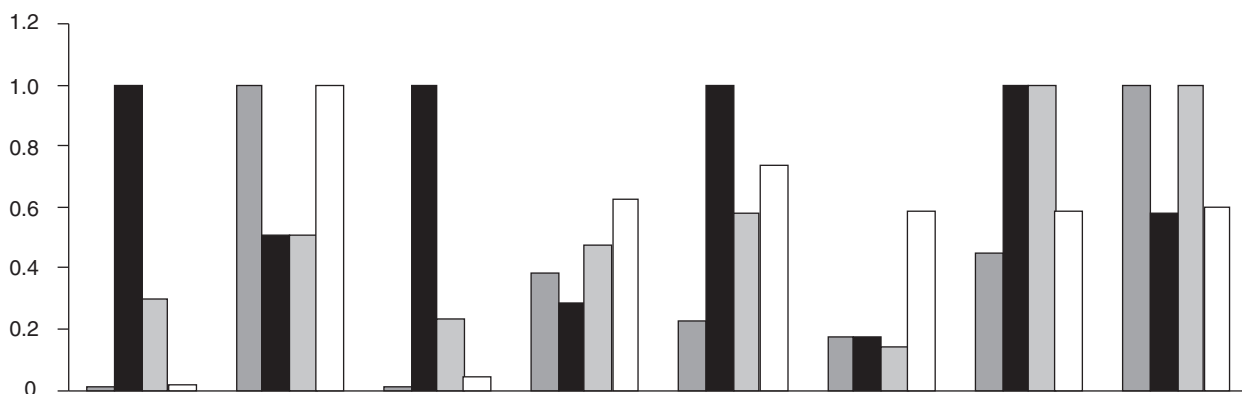
Regarding the narrative texts, differences occurred in five questions, namely four literal (two of microstructure and two of macrostructure) and only one inferential (macrostructure) when we analyzed group by group.

Each group answered 32 questions (eight for each of the four texts applied), therefore totaling 96 assessments.



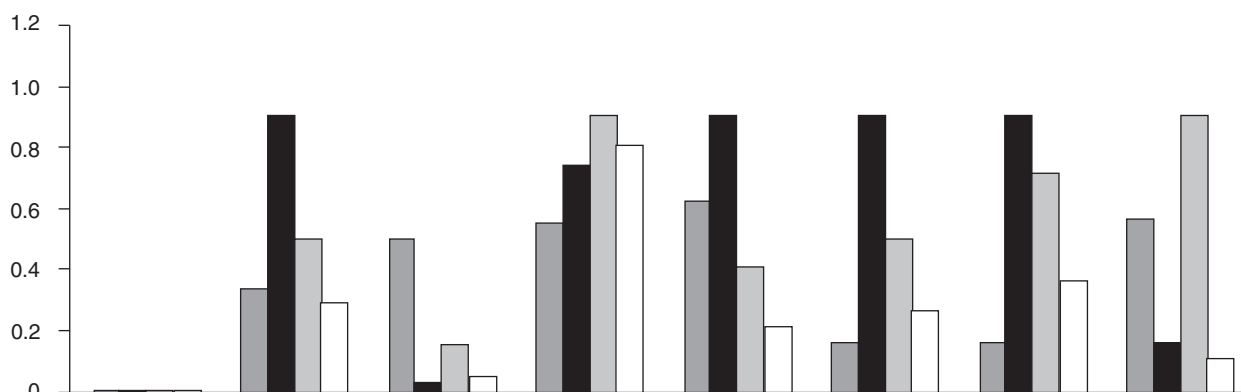
Statistical test used: likelihood ratio test

Graph 1. Average of the percentage of correct answers presented by the groups GI (third-year students), GII (fourth-year students), and GIII (fifth-year students) in the texts E1 (expository text 1), E2 (expository text 2), N1 (narrative text 1), and N2 (narrative text 2)



Statistical test used: McNemar's test

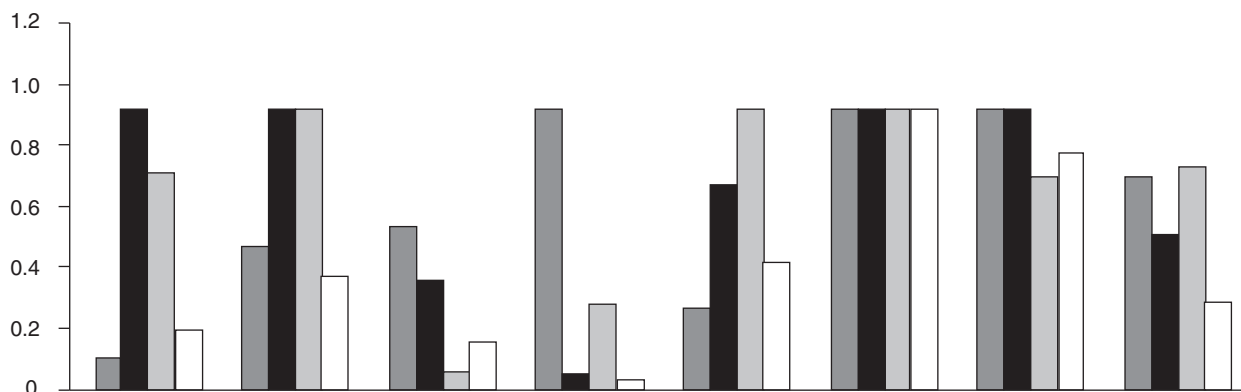
Graph 2. p-values obtained for each question (Q1 to Q8) upon comparison between the first and the second application of the expository text E1 for the groups GI (third-year students), GII (fourth-year students), and GIII (fifth-year students)



	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
■ GI	0.001	0.375	0.549	0.607	0.687	0.18	0.18	0.625
■ GII	0.001	0.999	0.031	0.815	0.999	0.999	0.999	0.18
■ GIII	0.001	0.549	0.167	0.999	0.453	0.549	0.791	0.999
□ Total	0.001	0.327	0.059	0.89	0.238	0.296	0.405	0.122

Statistical test used: McNemar's test

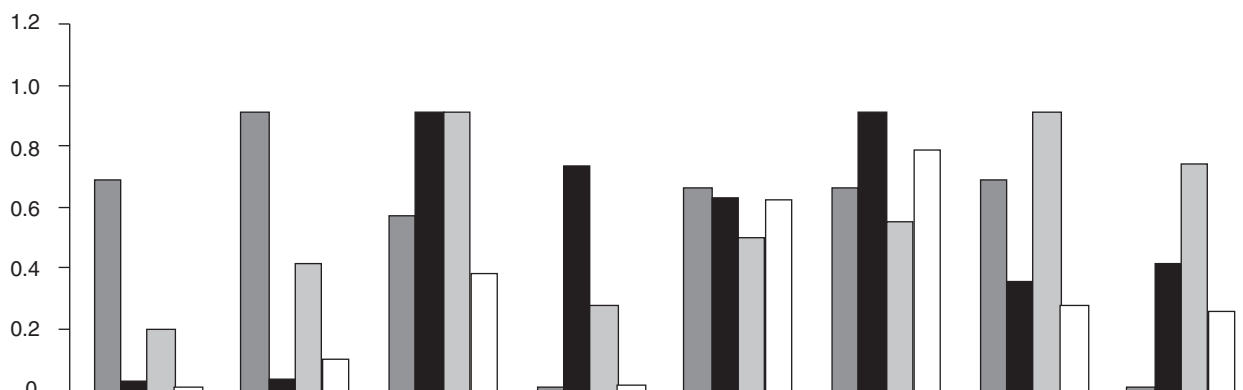
Graph 3. p-values obtained for each question (Q1 to Q8) upon comparison between the first and the second application of the expository text E2 for the groups GI (third-year students), GII (fourth-year students), and GIII (fifth-year students)



	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
■ GI	0.109	0.508	0.581	0.999	0.289	0.999	0.999	0.754
■ GII	0.999	0.999	0.388	0.057	0.727	0.999	0.999	0.549
■ GIII	0.774	0.999	0.064	0.302	0.999	0.999	0.754	0.791
□ Total	0.215	0.405	0.174	0.038	0.458	0.999	0.845	0.311

Statistical test used: McNemar's test

Graph 4. p-values obtained for each question (Q1 to Q8) upon comparison between the first and the second application of the narrative text N1 for the groups GI (third-year students), GII (fourth-year students), and GIII (fifth-year students)



	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
■ GI	0.1754	0.999	0.625	0.012	0.727	0.727	0.754	0.012
■ GII	0.031	0.039	0.999	0.804	0.687	0.999	0.388	0.453
■ GIII	0.219	0.454	0.999	0.302	0.549	0.607	0.999	0.815
□ Total	0.017	0.117	0.424	0.02	0.69	0.868	0.311	0.29

Statistical test used: McNemar's test

Graph 5. p-values obtained for each question (Q1 to Q8) upon comparison between the first and the second application of the narrative text N2 for the groups GI (third-year students), GII (fourth-year students), and GIII (fifth-year students)

In the analyses, we considered each group's performance in each question of the four texts, as well as the students' total sum. Thus, it was possible to verify through the results obtained that differences between the first and the second application of the instrument occurred in only 17 out of the 96 assessments, which amounts to a concordance of 82% between the two applications.

With the purpose of ascertaining and describing differences in the comparisons among the three groups studied with regard to the variables of interest, we conducted a statistical analysis using Kruskal-Wallis' test. In this analysis, we verified the incorrect answers provided by the students, with the purpose of assessing their performance in each question type (micro- and microstructural, literal and inferential).

Tables 2 and 3 display the data of each group with regard to each question type for the expository texts E1 and E2, respectively.

It is possible to verify from Tables 2 and 3 that statistically significant differences occurred in almost all variables for both expository texts, with the exception of macrostructural literal questions in the expository text E2 (Table 3). Concerning the expository text E1 (Table 2), upon comparing the averages among the variables, we verified that the average found for the macrostructural literal questions (maL) was higher than that of the microstructural literal questions (miL). The same occurred between micro- and macrostructural inferential questions (miI and maI), which suggests that

all groups had more difficulty with questions of macrostructure, both with literal and inferential questions.

Tables 4 and 5 display the data of each group with regard to each question type for the narrative texts N1 and N2, respectively.

Regarding the narrative texts, we verified that statistically significant differences also occurred in the majority of N1 variables (Table 4), thus indicating similar performances among the groups in this question type and text. For this text, the average of the groups in relation to the literal questions was higher when compared to the inferential questions. For the text N2 (Table 5), there were statistically significant differences in the majority of the variables as well. We also verified that the averages were higher for the N2 inferential questions.

DISCUSSION

One of the ways to assess the reliability of an instrument is to measure the concordance level between the answers obtained in two "moments of application" of the instrument for the same individuals. In this case, this level is expected to be high⁽⁹⁾, and this was verified through the results found in our analyses, which presented a concordance level higher than 82% between both applications for the same individuals.

Based on the evidence found through the answers analyzed statistically, we verified that concordance was high; therefore, we can consider that the same is true of reproducibility.

Table 2. Distribution of the average, standard deviation, and significance value found upon comparison among the groups for each variable in the expository text E1

Variable	n	Mean (SD)	p-value
E1 Lmi	102	0.83 (0.68)	<0.001*
I	121	0.76 (0.76)	
II	155	0.41 (0.61)	
III			
E1 Lma	102	0.86 (0.72)	0.022*
I	121	0.67 (0.71)	
II	155	0.62 (0.65)	
III			
E1 lmi	102	1.00 (0.77)	<0.001*
I	121	0.61 (0.74)	
II	155	0.46 (0.64)	
III			
E1 lma	102	1.12 (0.80)	<0.001*
I	121	0.66 (0.76)	
II	155	0.55 (0.77)	
III			

*Statistical test used: Kruskal-Wallis test.

Caption: E1 = expository text 1; miL = microstructural literal questions; maL = macrostructural literal questions; mil = microstructural inferential questions; mal = macrostructural inferential questions

Table 3. Distribution of the average, standard deviation, and significance value found upon comparison among the groups for each variable in the expository text E2

Variable	n	Mean (SD)	p-value
E2 Lmi	102	0.77 (0.83)	0.016*
I	121	0.62 (0.78)	
II	155	0.47 (0.66)	
III			
E2 Lma	102	0.95 (0.65)	0.588
I	121	0.97 (0.75)	
II	155	0.88 (0.69)	
III			
E2 lmi	102	0.88 (0.78)	0.002*
I	121	0.68 (0.71)	
II	155	0.56 (0.73)	
III			
E2 lma	102	1.12 (0.72)	<0.001*
I	121	0.89 (0.75)	
II	155	0.71 (0.77)	
III			

*Statistical test used: Kruskal-Wallis test

Caption: E2 = expository text 2; miL = microstructural literal questions; maL = macrostructural literal questions; mil = microstructural inferential questions; mal = macrostructural inferential questions

Table 4. Distribution of the average, standard deviation, and significance value found upon comparison among the groups for each variable in the narrative text N1

Variable	n	Mean (SD)	p-value
N1 Lmi	102	0.42 (0.70)	0.476
I	121	0.37 (0.65)	
II	155	0.32 (0.62)	
III			
N1 Lma	102	1.04 (0.77)	0.007*
I	121	0.82 (0.80)	
II	155	0.73 (0.72)	
III			
N1 lmi	102	0.91 (0.80)	<0.001*
I	121	0.67 (0.71)	
II	155	0.50 (0.67)	
III			
N1 lma	102	0.87 (0.75)	0.005*
I	121	0.89 (0.69)	
II	155	0.65 (0.70)	
III			

*Statistical test used: Kruskal-Wallis test.

Caption: N1 = narrative text 1; miL = microstructural literal questions; maL = macrostructural literal questions; mil = microstructural inferential questions; mal = macrostructural inferential questions.

Table 5. Distribution of the average, standard deviation, and significance value found upon comparison among the groups for each variable in the narrative text N2

Variable	n	Mean (SD)	p-value
N2 Lmi	102	0.56 (0.71)	0.953
I	121	0.55 (0.66)	
II	155	0.52 (0.61)	
III			
N2 Lma	102	0.70 (0.69)	<0.001*
I	121	0.46 (0.62)	
II	155	0.35 (0.55)	
III			
N2 lmi	102	0.87 (0.78)	0.074
I	121	0.76 (0.72)	
II	155	0.65 (0.70)	
III			
N2 lma	102	1.01 (0.78)	0.006*
I	121	0.87 (0.80)	
II	155	0.70 (0.71)	
III			

*Statistical test used: Kruskal-Wallis test

Caption: N2 = narrative text 2; miL = microstructural literal questions; maL = macrostructural literal questions; mil = microstructural inferential questions; mal = macrostructural inferential questions

Thus, it is possible to verify through the data shown that both “consistency” and “reproducibility” presented “high” results.

Therefore, we can ascertain that the instrument proposed in this study presented reliable foundations to be used in reading assessments, in accordance with previous arguments^{11,12} that a test with admittedly reliable and valid measurements provides researchers and clinical professionals with possibilities to adequately select the assessment tasks to be used in their work, as well as greater certainty in analyzing the data obtained. In this manner, these professionals are able to gather evidence that will support their scientific and clinical reasoning.

The differences verified suggest that some students behaved differently when dealing with propositions that contained explicit information and called for integration of textual propositions, requiring them to retrieve this integration from their work memory in order to be answered. They also presented differences of performance between the applications in relation to questions that required the elaboration of inferences among propositional elements, a necessary task in textual comprehension⁽⁴⁾.

In a longitudinal study⁽¹²⁾, scholars assessed whether work memory performance was related to reading comprehension improvement. The results showed that a relation occurred between these elements. Therefore, these results are in agreement with those found in our study, which also pointed out different performances concerning literal questions in the groups of individuals who are in later schooling stages; these questions require memory retrieval in order to be answered, thus influencing textual comprehension. Our data suggest, then, that the fifth-year students could employ their work memory development and their experiences with texts during the assessments, weighing on the answers given to the questions about comprehension of the texts applied.

Thus, the differences in performance found in the literal questions might have been influenced by memory and learning, considering that memory has the important role of evoking all knowledge acquired and stored (long-term memory) to be used in reading comprehension⁽¹³⁻¹⁵⁾.

We verified that the performance of the groups effectively improved starting from the third to the fifth year. This indicates that the students’ performances are quite different from one year to the other, which suggests that the reader’s interaction with the text occurs differently in the beginning of school life in comparison to more advanced years. In previous studies^(16,17), it was verified that students in the fourth, fifth and sixth years of elementary school had better reading comprehension performances than third-year students. Also, second-year students made more mistakes in reading comprehension than third-year students. With our results, we verified that the averages for the inferential questions were higher in the narrative texts, that is, the students made

more mistakes in this type of question. This corroborates a study⁽²⁾ in which it is reported that comprehension is a constructive and integrative process, insofar as able readers spontaneously make inferences in order to link ideas and obtain information that is only implicit; this is a necessary process for the elaboration of an integrated representation of the text. In this sense, students with reading comprehension problems may experience difficulties in making these inferences, as we have verified in our study.

The data presented here are also in agreement with previous studies^(18,19) in which scholars pointed out that the difficulty to make inferences limits the elaboration of an integrated representation of the meaning of a text, which, in turn, impairs comprehension.

Our data suggest that the expository texts were more difficult with regard to the questions that required memory in retaining information, as these were texts that contained specific information about a certain topic, whereas the narrative texts rely on a causal chain that organizes the events and actions that compose it, in addition to a temporal dimension⁽⁵⁾. Considering these aspects, the results of the present study indicate that the students had more difficulty to handle narrative elements when forming the inferences necessary to their comprehension.

Comprehension shortcomings create a nonspecific mental representation of the text that contains the general topic and a set of details linked to the theme. This means that the reader is unable to perceive the hierarchic relation around the ideas of a text, known as macrostructure, preventing him/her from linking the text to information that was acquired previously, therefore complicating the formation of inferences that are necessary for comprehension⁽²⁰⁾.

Thus, our results suggest that the students who had comprehension difficulties seem to experience this type of obstacle when forming the macrostructure of a text, which, in turn, interferes with inference elaboration. The data show, then, that, when answering the questions of a text, these students had difficulties to select the correct option, even though it was within their view. In other words, the answer did not seem clear to the students because they were unable to perceive the macrostructure that is necessary for inference elaboration.

CONCLUSION

Based on the data found, we conclude that the instrument elaborated proved to be efficient in its proposition to verify the reading comprehension profile of students, and to detect and characterize their difficulties. In this sense, the results point out that the performance differences presented by the students in the first application in relation to the second application of the comprehension assessment instrument, especially concerning literal questions, evidenced the role of memory in textual comprehension and its interference with inference elaboration;

the questions in which the students presented inferior performances were inferential and macrostructural.

The use of this reading assessment instrument can aid in developing specific interventions, with the purpose of helping students to overcome their problems. This is another way to prevent them from falling behind in comparison to their groups/classmates and to avoid the fact that this situation becomes an obstacle in their learning process and development.

Considering that this instrument was elaborated based on psychometric criteria, with data analyses that indicate its validity and internal consistency, we can conclude that it has reliable foundations that recommend its use to assess the reading comprehension of students from the third to the fifth year of elementary school.

ACKNOWLEDGMENTS

We thank

Coordination for the Improvement of Higher Education Personnel (CAPES) for granting a scholarship to the first author.

Prof. Emilio Sánchez Miguel and his research group at University of Salamanca (Spain), for collaborating with the construction of the instrument.

Adriana de Souza Batista Kida, PhD, for helping with text adaptation and the elaboration of the questions of the texts that compose the instrument.

**VLOC was responsible for data collection, tabulation and analysis, and manuscript writing; SAC was responsible for the project, study outline, and overall supervision of the stages of manuscript elaboration and writing.*

REFERENCES

- Alves ML. Identificação de níveis de compreensão de leitura – uma aproximação. *Enseñanza de las Ciencias*. In: VII Congreso Internacional. Universitat Autònoma de Barcelona; 2005. p. 1-5.
- Stothard S. Avaliação da compreensão da leitura. In: Snowling M, Stackhouse J, organizadores. *Dislexia, fala e linguagem*. São Paulo: Artmed; 2004. p. 121-42.
- Andrade MWCL, Dias MGBB. Processos que levam à compreensão de textos. *Psicol Est*. 2006;11(1):147-54.
- Kintsch W, Van Dijk TA. Toward a model of text comprehension and production. *Psychol Rev*. 1978;85(5):363-94.
- Miguel ES, Pérez JRG, Pardo JR. Leitura na sala de aula: como ajudar os professores a formar bons leitores. Porto Alegre: Penso; 2012.
- Cunha VLO, Oliveira AM, Capellini SA. Compreensão de leitura: princípios avaliativos e interventivos no contexto educacional. *Rev Teias*. 2010;10(19):1-21.
- Kida ASB, Chiari BM, Ávila CRB. Escala de leitura: proposta de avaliação das competências leitoras. *Rev Soc Bras Fonoaudiol*. 2010;15(4):546-53.
- Pinto AC. Factores relevantes na avaliação escolar por perguntas de escolha múltipla. *Psicol Educ Cult*. 2001;5(1):23-44.
- Pasquali L. Testes referentes a construto: teoria e modelo de construção. In: Pasquali L., organizador. *Instrumentos psicológicos: manual prático de elaboração*. Brasília: LabPAM/IBAP; 2001. p. 37-71.
- CFP: Conselho Federal de Psicologia [Internet]. Resolução nº 002/2003. Brasília: CFP; 2003 [citado 07 dez. 2010]. Disponível em: <http://www.pol.org.br>.
- Leslie L, Cadwell J. Formal and informal measures of reading comprehension. In: Israel SE, Duffy GG. *Handbook of research on reading comprehension*. New York: Taylor y Francis; 2009. p. 403-27.
- Lee SH. Dynamic testing, working memory, and reading comprehension growth in children with reading disabilities. *J Learn Disabil*. 2011;44(4):358-71.
- Kawano CE, Kida ASB, Carvalho CAF, Ávila CRB. Parâmetros de fluência e tipos de erros na leitura de escolares com indicação de dificuldades para ler e escrever. *Rev Soc Bras Fonoaudiol*. 2011; 16(1):9-18.
- Prior SM, Fenwick KD, Saunders KS, Ouellette R, O'Quinn C, Harvey S. Comprehension after oral and silent reading: does grade level matter? *Lit Res Instruc*. 2011;50(3):183-94.
- Cunha VLO, Silva C, Capellini SA. Correlação entre habilidades básicas de leitura e compreensão de leitura. *Estud Psicol*. 2012;29(Supl.): 799s-807s.
- Carvalho CAF, Ávila CRB, Chiari BM. Níveis de compreensão de leitura em escolares. *Pró-Fono R Atual Cient*. 2009;21(3):207-12.
- Lai MK, McNaughton S, Amituanai-Toloa M, Turner R, Hsiao S. Sustained acceleration of achievement in reading comprehension: The New Zealand experience. *Read Res Quarter*. 2009;44(1):30-56.
- Salles JF, Parente MAMP. Compreensão textual em alunos de segunda e terceira séries: uma abordagem cognitiva. *Estud Psicol*. 2004;9(1):71-80.
- Joly MCRA, Piovezan NM. Avaliação do Programa Informatizado de Leitura Estratégica para estudantes do ensino fundamental. *Paidéia*. 2012;22(51):83-90.
- Diakidoy IAN, Mouskounti T, Ioannides C. Comprehension and learning from refutation and expository texts. *Read Res Quarter*. 2011;46(1):22-38.

Appendix 1. Exemples of questions of the protocol**Exemplo de pergunta literal de microestrutura (texto expositivo E1)**

Há quanto tempo o homem é atingido pelo piolho?

- (a) Por quase três décadas.
- (b) Há cem anos.
- (c) Por dois séculos.
- (d) Há milhares de anos.

Exemplo de pergunta literal de macroestrutura (texto expositivo E2)

O que é a onça-pintada e o que acontece com ela?

- (a) É um animal que tem hábitos diurnos e que sobrevive em climas quentes caçando aves e pequenos animais.
- (b) É um felino que tem hábitos de caça noturnos e solitários e que se encontra à beira da extinção.
- (c) É um grande felino que se adapta em qualquer região climática e se encontra à beira da extinção.
- (d) É um animal que vive na América e sobrevive porque tem o hábito de caçar em bandos.

Exemplo de pergunta inferencial de microestrutura (texto narrativo N1)

Por que Henrique teve sorte ao chegar à escola?

- (a) Ele deixou o guarda-chuva escondido porque fazia sol.
- (b) Chegou atrasado, por isso conseguiu esconder o guarda-chuva.
- (c) Estava chovendo forte, por isso teve que usar seu guarda-chuva.
- (d) Só tinha um chuvisco, por isso deixou o guarda-chuva fechado.

Exemplo de pergunta inferencial de macroestrutura (texto narrativo N2)

Quando a professora perguntou de quem era o guarda-chuva, Henrique pensou que:

- (a) A professora ficaria brava porque deixou o guarda-chuva escondido.
- (b) Seus colegas também perguntariam de quem era o guarda-chuva.
- (c) Ele estava perdido e teria que falar que aquele guarda-chuva era dele.
- (d) A professora queria mostrar que o guarda-chuva era mesmo de Henrique.