

Escalas de avaliação da leitura e da escrita: evidências preliminares de confiabilidade****

Reading and writing assessment scales: preliminary reliability evidences

Adriana de Souza Batista Kida*
Brasília Maria Chiari**
Clara Regina Brandão de Ávila***

*Fonoaudióloga. Doutoranda em Distúrbios da Comunicação Humana pela Universidade Federal de São Paulo (Unifesp). Pesquisadora Associada do Núcleo de Ensino, Assistência e Pesquisa em Escrita e Leitura (Neapel) do Departamento de Fonoaudiologia da Unifesp. Endereço para correspondência: R. Vitorino Carmilo, 606 - Apto. 63 - São Paulo - SP CEP 01153-000 (adrianabatista@gmail.com).

**Fonoaudióloga. Livre-Docente pela Unifesp. Professora Titular da Disciplina dos Distúrbios da Comunicação Humana da Unifesp.

***Fonoaudióloga. Doutora em Distúrbios da Comunicação Humana pela Unifesp. Professora Associada do Departamento de Fonoaudiologia da Unifesp.

****Trabalho Realizado na Unifesp sendo parte da Tese de Doutorado apresentada ao Departamento de Fonoaudiologia da Unifesp - Escola Paulista de Medicina, para Obtenção do Título de Doutor em Ciências, com Fomento da Agência Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Artigo Original de Pesquisa

Artigo Submetido a Avaliação por Pares

Conflito de Interesse: não

Recebido em 01.12.2009.

Revisado em 06.10.2010.

Aceito para Publicação em 30.11.2010.

Abstract

Background: reliability of Reading and Writing Assessment Instruments. Aim: to investigate the reliability of two scales created to evaluate both reading and writing of children with ages between 8:0 and 11:11 years. Method: two scales were created: a reading scale, composed of 12 testing items organized into four competency fields (letter knowledge and phonographemic relation, decoding of isolated items, reading fluency, reading comprehension), and a writing scale, with five items organized into three fields (letter writing and graphophonemic relation, codification of isolated items, writing construction). One hundred students (64 girls) from Public Schools, with ages ranging between 8:0 and 11:11 years, were selected. Twenty students (12 girls) participated in the applicability study, resulting in the study version of the Scales. These scales were later applied to the remaining 80 students (52 girls). The obtained responses were assessed and computed for score assignment: item scores, competence field score (CFS) and raw scale score (RSS). Data were submitted to statistical analysis: the Cronbach's alpha coefficient was calculated and correlations between items (Pearson's correlation coefficient) were verified. A significance level of 0.05 was used. Results: $\alpha = 0.866$ and $\alpha = 0.461$ were obtained for the Reading and Writing Scales, respectively. Correlations between the items were observed, ranging from weak to strong, and confirmed the alpha values. Conclusion: the Reading Scale was proven reliable, achieving acceptable levels for diagnostic instruments; the Writing Scale did not present an acceptable reliability level to measure the performance of the tested children.

Key Words: Reading; Handwriting; Assessment; Speech; Language and Hearing Sciences.

Resumo

Tema: confiabilidade de Instrumentos de avaliação da leitura e escrita. Objetivo: investigar a confiabilidade de duas escalas elaboradas para a avaliação da leitura e escrita de crianças de 08 a 11:11 anos. Método: foram elaboradas duas escalas: de leitura, composta por doze itens de testes organizados em quatro campos de competências (conhecimento de letras e relação fono-grafêmica, decodificação de itens isolados, fluência de leitura de textos, compreensão de leitura), e de escrita com cinco itens organizados em três campos (escrita de letras e relação grafo-fonêmica, codificação de itens isolados, construção escrita). Selecionaram-se 100 escolares (64 meninas) de rede pública com idade de 8 a 11:11 anos. Vinte (12 meninas) participaram do estudo de aplicabilidade, que resultou na versão de estudo das escalas, posteriormente aplicadas aos demais 80 escolares (52 meninas). As respostas obtidas foram analisadas e computadas para atribuição dos escores de itens, escores por campo de competência (ECC) e do escore bruto da escala (EBE). Os dados foram analisados estatisticamente, obtidos o coeficiente alpha de Cronbach e, complementarmente, as correlações entre os itens (coeficiente de correlação de Pearson). Adotou-se nível de significância de 0,05. Resultados: Obtiveram-se $\alpha = 0,866$ e $\alpha = 0,461$ para as escalas de leitura e escrita, respectivamente. Correlações entre os itens foram observadas, variando de fracas a fortes e corroboraram os valores de alpha. Conclusão: a escala de leitura mostrou-se confiável, atingindo níveis admissíveis para instrumentos diagnósticos, enquanto que a escala de escrita não apresentou nível de confiabilidade admissível para mensurar o desempenho das crianças da amostra.

Palavras-Chave: Leitura; Escrita; Avaliação; Fonoaudiologia.

Referenciar este material como:



Kida ASB, Chiari BM, Ávila CRB. Reading and writing assessment scales: preliminary reliability evidences (original title: Escalas de avaliação da leitura e da escrita: evidências preliminares de confiabilidade). Pró-Fono Revista de Atualização Científica. 2010 out-dez;22(4):509-14.

Introduction

Like any other disorder, clinical identification of Reading and Writing Disorders needs valid, reliable, standardized and normalized instruments that are capable to support the diagnosis, definition of conduct and organization of intervention programs¹. Protocols that do not meet these specifications may undermine the reliability of the clinical evidences needed for diagnosis². Among the national tests available³⁻⁸, some evaluate aspects of underlying processes to reading and writing skills. However, these do not present reference norms and do not enable the comparison of the assessed performances. Other tests are restricted and specifically assess the recognition and decoding skills in reading and encoding of single items in writing during dictation without presenting any reference norms and reliability data. There are also tests with reference norms which, however, have no reliability data.

This study aimed to investigate the reliability of two scales designed to assess reading and writing of children from eight years to 11y11m, assuming the hypothesis that both scales of measurement would present characteristics of diagnostic instruments. To illustrate the importance of studying the measurement properties of the material proposed for diagnostic tests, the scales and particularly the method adopted for the preparation and reliability analysis of these instruments will be presented.

Method

This study was approved by the IRB/UNIFESP (number 11.111/05). The study was divided into six stages, according to the standards for construction of cognitive evaluation tests¹: development of scales, sample selection, preliminary study of applicability; application of scales and collection of responses; attribution of scores; study of internal consistency.

The elaboration of Reading Scales (RS) and Writing Scales (WS) considered the following rules: the defined objectives; the literature review of assessment methods; the definition of the test format; and the selection of assessment materials, test stimuli and analysis procedures.

A priori, it was decided that the scales would evaluate the performance of students from eight to 11y11m regarding the skills involved in learning and reading and writing capabilities. Following, indicators of reading 2, 9-11 and writing 12-15 performance appropriate to the defined age range, tests and assessment methods of reading 4; 16-20 and writing 3-

4, 16, 21-23 were analyzed. From the abilities to be investigated, the following fields were delimited: knowledge of letters and of the graph-phonemic relationship; decoding of isolated items; text reading fluency and comprehension for RS; and knowledge of the phono-grapheme relationship; coding; and construction of writing for WS.

Test items were defined for each scale and were selected: the linguistic material, the assessment procedures (including application instructions), and criteria for performance analysis for each item.

In the preliminary version, the RS contained twelve test items and the WS, five (Table 1).

Participants were 100 schoolchildren (64 girls) between eight and 11y11m of age, from Second to Sixth grade of Public Elementary Schools. Participants were selected from 132 children nominated by teachers as having good academic performance. The selection of participants was determined according to the following inclusion criteria: absence of complaints or indicators of hearing and/or visual deficits; no neurological, behavioral or cognitive impairments; no complaints of difficulties or learning disabilities; no complaints of school performance; absence of retention in school history; signature of the consent form; passing the Speech, Language and Hearing screening^{6;24-25}: tests of verbal and writing production and comprehension. Participants were divided into subgroups according to age: 8-8y11m; 9-9y11m; 10-10y11m; 11-11y11m. Five children of each subgroup were randomly selected to compose the sample of the Preliminary Applicability Study¹ that evaluated the first version of the Scales and indicated the need for modifications. Scales were applied on the 20 children for observation of: duration of application of the test scales, effectiveness of instruction comprehension, difficulty degree of test stimuli proposed according to age group, and the functionality of Performance Record Sheets (FRD). Two test items from the WS and FRD of both scales were modified.

After the final version was determined, each of the 80 participants was evaluated during two sessions (35 minutes each) in school rooms with noise levels that did not interfere on the comprehension of the instructions.

We adopted a standardized scoring system to analyze the performance and enable the reliability study: scores per item, per field of competence (ecc) and total score per scale (cee).

The scores of items ranged from 2 to 0 and represented, respectively, better to poorer performance. The performances, quantified according to the criteria for analysis (Table 1), were tabulated for each child and statistically treated. Measures of central tendency were calculated and the median and the third

quartile were adopted as parameters. Score attribution criteria were defined for items according to each age range. The median represents score 2, values between the median and the third quartile represent score 1, and values below the third quartile represent the score zero.

The ecc (sum of scores per item) reported the performance on each Competence Field and allowed to consider whether the items in each field examined a single competence. The cee (sum of ecc) reported the overall performance on the scales. It allowed verifying whether all the selected items were related to the construct of reading or writing.

The study of internal consistency of the scales fulfilled the purpose of analyzing the reliability; 26 different test items should measure the same variable 26. The inter-items internal consistency was analyzed 26 28. The responses of each participant were reevaluated and received a score (score per item). The other scores (ecc and ebe) were also computed for each participant.

Results

The Statistical Analysis System version 13.0 was used for the analysis. The Cronbach's alpha coefficient (α) was applied to analyze the internal consistency by studying the degree of covariance of the items. The congruence of each item of a scale with other items that comprised the scale, and the effect of each item on the instrument, were verified through additional study of reliability after the removal of test items. Values of α below 0.6 indicated degree of covariance at unacceptable levels; values between 0.6 and 0.7 indicated weak

covariance; values between 0.7 to 0.8 indicated acceptable covariance; values 0.8 and 0.9 indicated good covariance degree; and values above 0.9 indicate very good covariance.

The level of significance for this study was 0.05.

The internal consistency of RS and its fields of competence (Table 1) exhibited a good degree of covariance for the fields of decoding isolated items, text reading fluency, and for the total scale. The field of reading comprehension showed very low degree of covariance.

The suppression of these items revealed that the α values maintained the good covariance to decode isolated items (deletion of item 3: $\alpha = 0.848$; item 4: $\alpha = 0.855$; item 5: $\alpha = 0.870$; item 6: $\alpha = 0.896$), text reading fluency (deletion of item 7: $\alpha = 0.855$; item 8: $\alpha = 0.812$; item 9: $\alpha = 0.893$; item 10: $\alpha = 0.848$), and for the total RS (deletion of item 2: $\alpha = 0.887$; item 3: $\alpha = 0.843$; item 4: $\alpha = 0.840$; item 5: $\alpha = 0.846$; item 6: $\alpha = 0.848$; item 7: $\alpha = 0.842$; item 8: $\alpha = 0.835$; item 9: $\alpha = 0.853$; item 10: $\alpha = 0.838$; item 11: $\alpha = 0.882$; item 12: $\alpha = 0.871$). Because the fields of knowledge of letter and reading comprehension consisted of only two items, their internal consistency from the removal of items could not be analyzed.

The internal consistency of WS and its fields of competence (Table 2) revealed very low degrees of covariance. The internal consistency of writing construct was not analyzed because it is constituted by a single test item.

The exclusion of items 1 and 2 increased the covariance of this scale, which, however, did not increase the coefficient of WS to acceptable levels maintaining the very low variance ($\alpha = 0.506$).

FIGURE 1. Items and Analysis Criteria of the Reading and Writing Scales

Scale	Test items	Assessment			
		Objective	Material	Criteria	
RS	1 and 2	Knowledge of letters and their respective and possible phonemes		> All the letters of the alphabet except K, W and Y for the item 1 ¹⁶⁻¹⁷ > All the letters of the alphabet except for H, K, W, Y for item 2 ¹⁶ > accuracy of correct naming of letters ¹⁶⁻¹⁷ > accuracy on the correct assignment of the sound value of letters ¹⁶	
	3 and 4	Decoding – reading of isolated items	words	> 38 revised items ¹⁶ balanced for: extension, familiarity e orthography.	
	5 and 6		pseudo-words	> 29 items ¹⁶ balanced for: extension familiarity and orthography.	
	7, 8, 9 and 10	Text reading fluency		> 4 texts (selected from didactic materials from each grade)	> parameters of rate and reading accuracy ¹⁰⁻¹¹ in situations of first and second reading
	11 and 12	Text comprehension		> 4 other texts: "The monkey and the rabbit" - Monteiro Lobato for eight year-olds and other texts adopted for the other age ranges ²¹	> recalling of text read analysis ¹⁹ : identification of the number of central ideas and retold causal relationships > accuracy to 6 multiple-choice questions ²⁰ : refer to explicit propositions (short-term memory) or implicit information from text (inferential comprehension)
WS	1 and 2	Writing of letters from their name and production of their sound.		> All the letters of the alphabet except K, W and Y for item 116 > All the letters of the alphabet except for H, K, W, Y for item 216 > accuracy ¹⁶	
	3	Coding – writing of dictated isolated items	words	> 35 words balanced according to: orthography and familiarity presupposed by the EF board of teachers and extension – two- and three-syllable ¹⁶ > accuracy ¹⁶	
	4		pseudo-words	> 21 revised pseudo-words ¹⁶ > accuracy ¹⁶	
	5	Written elaboration from the problem situation orally proposed by the examiner for the sequence of figures		> Test cards as visual support material: selected according to determinations of the National Curriculum ²² > Scoring system based on appropriateness of the use of punctuation, discursive coherence and cohesion ²²⁻²³	

Note: RS= Reading Scale; WS= Writing Scale

TABLE 1. Measures of internal consistency of RS and its competence fields

Skills and Competence Fields	Test items	α
Letter knowledge	1 to 2	Not calculated *
Codification of isolated items	3 to 6	0,898
Text reading fluency	7 to 10	0,887
Reading comprehension	11 to 12	0,317
Total scale	1 to 12	0,866

* α was not calculated because values of item 1 are constant

Note: α = Cronbach's alpha coefficient

TABLE 2. Measures of internal consistency of WS and its competence fields

Skills and Competence Fields	Test items	α
Letter knowledge	1 and 2	0,070
Codification of isolated items	3 and 4	0,560
Total Scale	1 and 5	0,461

Note: α = Cronbach's alpha coefficient

Discussion

The remarkable importance of tests and assessment procedures in the clinical practice indicates that cognitive testing should have characteristics that demonstrate its measurement properties: validity and reliability. Tests and assessment procedures should provide accurate and stable performance in a particular skill; sensitivity and specificity in appropriately identifying healthy subjects and patients with alterations; and normative guidelines that are essential to the diagnosis¹⁻², 26-27.

In order to meet these determinations, the elaboration of the Reading and Writing Scales followed the standards of tests construct¹. The reliability study showed that the different selected tasks did not always examine the same pretended content or processing, indicating that selection of new assessment procedures should be performed. This would increase the security in collecting and analyzing the evidences that support scientific and clinical reasoning².

The study of inter-items internal consistency - adopted as a form of evaluation of the scales - was performed to answer whether the selected items were related to an unique theoretical construct (sometimes understood as Reading sometimes as Writing) represented by ebe of each scale, or even like each one of the Fields of Expertise, represented by ecc. The presence of admissible values of covariance - higher than 0.728 - was required to certify the reliability of the instruments on the performance measurement properties according to age.

The selection of children from the Public School System of only one region (São Paulo) aimed at the same socio-cultural profile. The indication of the best schoolchildren by the teachers tried to minimize the influence of possible effects of language disorders and learning disabilities on the evaluation of the capability of measuring the instruments.

The adequacy of the RS for diagnosis was validated by the internal consistency of the Scale and of the isolated items of fields of decoding and text reading fluency. The selected items measured the same construct, demonstrating its reliability^{26, 28-29}. The field of knowledge of letters, for which α was not calculated, showed a ceiling effect that prevented the analysis of internal consistency.

The field of reading comprehension was the only one that did not reach levels of diagnostic reliability. However, the exclusion of items 11 and 12 did not alter the consistency of the RS, suggesting that both contribute to the assessment of reading capacity. The use of different assessment procedures may require different skills and abilities to respond to the test²⁶, possibly interfering on the internal consistency values. Thus, the low values of α for the Field of comprehension may have been determined by the variability of the demands imposed by the evaluation procedures²: retelling the text read and answering multiple choice questions.

The internal consistency of WS revealed inadequate instrument for diagnostic²⁹ or screening²⁶. Thus, the theoretical concept that guided the construction of the WS was not properly represented by the items selected to compose this instrument. Two factors may have interfered: the heterogeneity of assessment procedures of test items and the absence of validity data attesting the reliability of the items selected from the literature to assess a particular construct.

The present research demonstrated the importance of the study and adoption of reliability parameters in the construction of instruments of clinical assessment. Moreover, the data indicated that RS must undergo further testing on its discriminating property, sensitivity and specificity to certify its accuracy for clinical use. Subsequently, one should proceed to the study of

normalization, which will provide performance parameters of typically developing children according to age and education, diagnostic criteria established by DSM IV30. As for WS, items should be revised and the reliability restudied. Only these clinical studies will attest to the clinic the quality of the instruments in obtaining reliable data for diagnosis and therapeutic planning in Reading and Writing.

Conclusion

The reliability study of the Reading and Writing Scales indicated the possibility of using the RS to evaluate and diagnose competences related to letter knowledge and phono-grapheme relationship, decoding, and reading comprehension. In contrast, WS proved to be inadequate for measuring the performance of schoolchildren related to the knowledge of phono-grapheme relationship, coding, and construction of writing, requiring substantial revisions and re-evaluation of its measurement properties.

References

1. American Psychology Association (APA). Standards for educational and psychological testing. New York: American Educational Research Association, 1999.
2. Leslie L, Caldwell JA. Formal and Informal measures of reading comprehension. In: Israel SE, Duffy GG, editores. Handbook of research on reading comprehension. New York: Routledge; 2009. p.403-427.
3. Braz HA, Pellicciotti THF. Exame de linguagem Tipiti. São Paulo: MNJ LTDA; 1988.
4. Stein LM. Teste de Desempenho Escolar: Manual de aplicação e interpretação. São Paulo: Editora Casa do Psicólogo, 1994.
5. Capovilla FC, Capovilla AGS. Problemas de leitura e escrita: como identificar, prevenir e remediar numa abordagem fônica. Ed. Memnon, São Paulo: 2000.
6. Seliar-Cabral L. Guia prático de alfabetização, baseado em princípios alfabéticos do português do Brasil. São Paulo: Ed. Contexto, 2003.
7. Saraiva RA, Moojen SMP e Murarki R. Avaliação da compreensão leitora de textos expositivos. São Paulo: Casa do Psicólogo, 2006.
8. Capellini SA, Cunha VLO. Provas de habilidades metalingüísticas e de leitura. São Paulo: Revinter, 2009.
9. Seymour PHK. Individual cognitive analysis of competent and impaired reading. Br J Psychol. 1987;78:483-506.
10. Jenkins J, Fuchs L, Fuchs D, Hosp M. Oral reading competence: a theoretical, empirical and historical analysis. Sci Stud Read. 2001;5(3):239-56.
11. Geva E, Zadeh ZY. Reading efficiency in native English-speaking and English-as-a-second-language children: the role of oral proficiency and underlying cognitive-linguistic processes. Sci Stud Read. 2006;10(1):31-57.
12. Bishop DVM, Clarkson B. Written language as a window into residual language deficits: a study of children with persistent and residual speech and language impairments. J.cortex. 2003;39:215-37.
13. Puranik CS, Lombardino LJ, Altmann LJ. Writing through retellings: an exploratory study of language-impaired and dyslexic populations. Read Writ. 2007;20:251-72.
14. Puranik CS, Lombardino LJ, Altmann LJP. Assessing the microstructure of written language using a retelling paradigm. Am J Speech Lang Pathol. 2008;17:107-20.
15. Caravolas M, Hulme C, Snowling MJ. The foundations of spelling ability: evidence from a 3-year longitudinal study. Journal of Memory and Language. 2001;45:751-74.
16. Ramos CS. Avaliação da leitura em escolares com indicação de dificuldades de leitura e escrita [tese]. São Paulo(SP): Universidade Federal de São Paulo; 2005.
17. Cuetos FV. Psicología de la lectura: diagnóstico y tratamiento. Madrid: Ed. Escuela Espanola; 1990.
18. Goikoetxea E. Reading errors in first and second-grade readers of a shallow orthography: evidence from Spanish. British Journal of Educational Psychology. 2006;76:333-50.
19. Trabasso T, van der Broek P. Causal thinking and the representation of narrative events. Journal of memory and language. 1985;24:617-30.
20. Carvalho CAF, Ávila CRB, Chiari BM. Níveis de compreensão de leitura em escolares. Pró-Fono R Atual Cient. 2009 jul-set;21(3):207-12.
21. Bereiter C, Burtis PJ, Scardamalia M. Cognitive Operations in constructing main points in written composition. Journal of memory and language. 1988(27):261-78.
22. Ministério da Educação (Brasil). Secretaria de Educação Fundamental. Parâmetros curriculares nacionais: língua portuguesa. Brasília: Secretaria de Educação Fundamental; 1997.
23. Gelderen AV, Oostdam R. Effects of fluency training on the application of linguistic operations in writing. Educational Studies in Language and Literature. 2005;5:215-40.

24. Sanchez Miguel E. Compreensão e Redação de Textos: dificuldades e ajudas. Porto Alegre: Ed. Artmed, 2002.
25. Paolucci JF, Avila CRB. Competência ortográfica e metafonológica: influências e correlações na leitura e escrita de escolares da 4a série. Rev Soc Bras Fonoaudiol. 2009;14 (1):48-55.
26. Domino G, Domino ML. Psychological Testing. New York: Cambridge University Press; 2006.
27. Goulart BNG, Chiari BM. Testes de rastreamento x testes de diagnóstico: atualidades no contexto da atuação fonoaudiológica. Pró-Fono R Atual. 2007;19(2):223-32.
28. Messick S. Validity of psychological assessment. American Psychologist. 1995;50:741-49.
29. Dahlstrom WG. Tests. Small samples, large consequences. American Psychologist. 1993;48:393-9.
30. American Psychology Association (APA). Diagnostic and Statistical Manual of Mental Disorders - 4th. Washington: American Psychology Association; 1994.