

Morphosyntactic Evaluation Protocol (MEP):
validation of content*Protocolo de Avaliação Morfossintática (PAM):
validação de conteúdo*Lorena Adami da Cruz Pinheiro¹ Anilton Pinheiro da Silva² Simone Rocha de Vasconcellos Hage³ 

Keywords

Language Child
Language Development
Language Tests
Developmental Language Disorder
Validation Studies

Descritores

Linguagem Infantil
Desenvolvimento da linguagem
Testes de Linguagem
Transtorno do Desenvolvimento da
Linguagem
Estudos de Validação

Correspondence address:

Simone Rocha de Vasconcellos
Hage. Universidade de São Paulo –
Departamento de Fonoaudiologia – FOB/
USP, Alameda Dr. Octávio Pinheiro
Brisolla 9-75 – Vila Universitária, Bauru,
SP, Brasil.
E-mail: simonehage@usp.br

Received: July 14, 2019.

Accepted: February 03, 2020.

ABSTRACT

Introduction: the use of language assessment instruments in the area of speech-language-therapy is essential for the diagnosis and, consequently, for therapeutic planning. In Brazil, there is a shortage of instruments constructed and validated in the morphosyntax area. Morphosyntactic Evaluation Protocol (MEP) was constructed based on the main syntactic characteristics of the period of acquisition of children's language, on the Portuguese grammatical structure and the application in a pilot study. **Objective:** To verify the validity of MEP content. **Methods:** for the validation process, the instrument was applied and analyzed through a questionnaire by three judges, a linguist and two speech-language specialists with experience in assisting children with Language Development Disorder. The Index of Judges' Reliability was used to compare the results of the protocol application and the Cronbach's Alpha tests, Spearman-Brown and Content Validity Index (CVI) in the questionnaire responses. **Results:** the statistical tests applied in the validation of content legitimized the reliability of the instrument with indexes considered substantial for both alpha coefficients, higher than 0.80, Spearman and the CVI test had a maximum concept of 1.0. **Conclusion:** there were compliance and compatibility in the answers of the experts, which indicates the reliability of the instrument. The results of the statistical tests legitimize the reliability of the instrument with indexes considered substantial for alpha and Spearman coefficient. In the future, the protocol may help characterize the syntactic profile of children with language developmental disorder.

RESUMO

Introdução: a utilização de instrumentos de avaliação de linguagem na área da Fonoaudiologia é essencial para a realização do diagnóstico e, conseqüentemente, para o planejamento terapêutico. No Brasil há escassez de instrumentos construídos e validados na área da morfossintaxe. Protocolo de Avaliação Morfossintática (PAM) foi construído com base nas principais características sintáticas do período de aquisição da linguagem infantil, na estrutura gramatical da língua portuguesa e na aplicação em estudo piloto. **Objetivo:** verificar a validade de conteúdo do PAM. **Método:** para o processo de validação, o instrumento foi aplicado e analisado por meio de questionário por três juízes, um linguista e dois fonoaudiólogos especialistas em linguagem com experiência no atendimento de crianças com transtorno de linguagem. Foi usado o Índice de Fidedignidade entre os juízes para comparação dos resultados da aplicação do protocolo e os testes Alpha de Cronbach, Spearman-Brown e Índice de Validade de Conteúdo (IVC) nas respostas do questionário. **Resultados:** os testes estatísticos aplicados na validação de conteúdo legitimaram a fidedignidade do instrumento com índices considerados substanciais tanto para coeficiente alpha, maior que 0,80, Spearman e o teste IVC teve conceito máximo de 1,0. **Conclusão:** houve conformidade e compatibilidade nas respostas dos especialistas o que indica confiabilidade do instrumento. Os resultados dos testes estatísticos legitimam a fidedignidade com índices considerados substanciais tanto para coeficiente Alpha, como Spearman. No futuro, o protocolo poderá auxiliar na caracterização do perfil sintático de crianças com alterações no desenvolvimento de linguagem.

Study conducted at the Universidade de São Paulo – Departamento de Fonoaudiologia – FOB/USP. Bauru (SP), Brasil.

¹ Programa de Pós-Graduação em Fonoaudiologia – Linha de Pesquisa: Processos e Distúrbios da Linguagem, Bauru (SP), Brasil.

² Programa de Pós-Graduação em Fonoaudiologia – Linha de Pesquisa: Processos e Distúrbios da Audição, Bauru (SP), Brasil.

³ Universidade de São Paulo – Departamento de Fonoaudiologia – FOB/USP Bauru (SP), Brasil.

Conflict of interest: nothing to declare.

Financial support: FAPESP – Regular Programs/Regular Research Project – 2017/02990-9.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Morphosyntactic knowledge is decisive for the child to evolve from the level of isolated words to that of utterances using grammatical rules of the language. When the child starts to establish relations between the words in a sentence and use crunchees around the age of 18 months, morphosyntax is established⁽¹⁾. In the clinical context, this knowledge is commonly affected in Language Disorders, whether they are primary or not⁽²⁾. Despite the importance that morphosyntax has in the development of child communication, it has been neglected in Brazilian studies because it is not possible to measure it reliably in Brazilian Portuguese. Regarding the specific verification of the morphological and syntactic structure of the language, only the MLU - Mean Length Utterance has been used as a measurement procedure in recent years^(3,4). In this context, the construction or cross-cultural adaptation of syntactic assessment instruments represents a challenge for advances in the evaluation of children's language. To overcome this challenge, the instruments must be validated for Brazilian speakers.

To contribute to the morphosyntactic evaluation of children whose language is Brazilian Portuguese, the morphosyntactic evaluation protocol (MEP) was developed by Brazilian researchers and involved two stages: first, categories were listed in the evolutionary models of morphosyntactic acquisition based on literature review. The review sought information from different bibliographic sources on the morphosyntactic evaluation of children and the grammar of the Portuguese language. The implicit rules of language use and the language acquisition phases of children between 2.6 and 5.0, age group covered by the protocol, were considered. In a second step, the protocol was applied in a pilot group consisting of both children with typical language development, as well as with primary Language Disorder, that is, not associated with a biomedical condition. This application was essential to verify whether the categories identified in the literature were compatible with the speech corpus of children with the disorder. Based on the grammatical structure of the Portuguese language and the most frequent characteristics of the morphosyntactic construction of children with language disorders, MEP initially chose eight analysis criteria to be verified on the orthographic conversation transcript between child and adult interlocutor for 20 minutes. The analysis criteria were: 1. Number of statements produced, 2. Number of ungrammatical phrases, 3. Number of telegraphic phrases, 4. Number of noun phrases, 5. Number of simple periods, 6. Number of compound periods, 7. Number of errors of nominal agreement and 8. Number of errors of verbal agreement.

However, for an assessment instrument created to be used reliably, it is essential to check its psychometric qualities. The importance of verifying these qualities in language assessment procedures is increasingly emphasized⁽⁵⁾, and one of these qualities is the validity in which it is verified whether an instrument measures precisely what it proposes to measure⁽⁶⁾. There are

several ways to validate an instrument, one of which is content. Content validity refers to the degree to which the content of an instrument reflects the construct that is being measured and implies an evaluation with quantification of judgments, whose main objective is to evaluate and improve the criteria used in an instrument⁽⁷⁾.

This study aimed to verify the content validity of an instrument created by Brazilians to verify the morphosyntactic abilities of children in language development.

METHODS

The research was approved by the Human Research Ethics Committee of "Faculdade de Odontologia de Bauru" (CAAE 68562317.4.0000.5417). The judges signed the Free and Informed Consent Form (FICF), according to Resolution 466/2012.

Studies on content validation are divergent concerning the number of reviewers. The recommendation can vary from two to twenty⁽⁷⁾. Lynn⁽⁸⁾ points out that the minimum number of judges can be three. Thus, the choice was made for three expert evaluators with experience in evaluating all aspects covered in the protocol. That said, we invited three experts in the language field, two speech-language therapists and a linguist who were all willing to apply to five speech samples and evaluate the instrument.

Content validation was carried out in two stages. First, the judges received a transcript of the speech sample of five children with Language Disorder, as their speech was the ultimate goal of creating the protocol, and applied the protocol based on the guidelines provided through a "step by step" manual. After the application, the judges evaluated the instrument through a questionnaire with questions regarding the pertinence, relevance, applicability and representativeness of each criterion of the protocol. The questions about each item served to assist in the judgment and the judges' evaluation. The use of a questionnaire to obtain expert judgment is a method that minimizes bias and standardizes the requested information on the content of each item⁽⁷⁾.

For the statistical analysis of the questionnaire, the Likert scale⁽⁹⁾ with four levels of support was used: A - Disagree; B - Partially disagree; C - Partially agree; D - Agree.

With data from the application of the protocol by the judges, a table was generated to which the calculation of agreement between them or the reliability index was applied. The calculation was performed using the technique in which the reliability index must be equal to or greater than 70%⁽¹⁰⁾, which recommends a fact not to be produced at random, therefore indicating reliability.

Cronbach's test⁽¹¹⁾ was applied to each judge's questionnaire response list, a statistical measure that ensures the internal consistency of a test or scale. Cronbach (11) quantified this reliability by proposing a coefficient, α , which varies from 0 to 1. If α is close to 0, then the qualified answers are not reliable, and if it is close to 1, the answers are very reliable. If $\alpha \geq 0.8$, then the answers are considered reliable.

The reliability of the instrument was also verified using the halves method, that is, the items that make up the MEP were grouped into two halves, comparing the scores obtained to these halves. Precision coefficients were estimated using the Spearman-Brown formula⁽¹²⁾.

RESULTS

Table 1 represents the data obtained through the application of the protocol by the judges in a speech sample already transcribed from five children with Language Disorder. The analysis of the responses was based on eight protocol criteria.

Table 1. Reliability Index among the judges on the application of the protocol in a spontaneous speech sample of five children

MEP criteria	Reliability Index (RI) among judges on the application of the protocol on spontaneous speech samples														
	Reliability Index					Reliability Index					Reliability Index				
	Judges 1x2					Judges 1x3					Judges 2x3				
	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
1	0.94	0.97	0.99	0.80	0.99	0.97	0.96	0.95	0.80	0.99	0.97	0.99	0.96	1	0.98
2	1	0.83	0.77	1	1	1	1	1	1	0.80	1	0.83	0.77	1	0.80
3	1	1	1	1	0.95	1	1	1	1	1	1	1	1	1	0.95
4	0.95	0.95	0.92	1	0.93	1	0.95	1	1	0.96	0.95	1	0.92	1	0.96
5	0.95	0.98	0.96	0.92	0.94	1	0.94	0.97	0.91	0.99	0.95	0.93	0.99	0.83	0.93
6	1	0.89	0.88	1	0.92	0.88	0.88	0.88	1	0.85	0.88	1	1	1	0.92
7	1	1	0.83	1	0.83	1	1	1	1	1	1	1	0.83	1	0.83
8	0.80	1	1	1	0.80	0.80	1	1	1	1	1	1	1	1	0.80

Captions. C = child. 1. Number of statements produced, 2. Number of ungrammatical phrases, 3. Number of telegraphic phrases, 4. Number of nominal phrases, 5. Number of simple periods, 6. Number of compound periods, 7. Number of errors of nominal agreement and 8. Number of verbal agreement errors.

Table 2 refers to the results obtained with the application of the Cronbach's *Alpha*⁽¹¹⁾, Spearman-Brown⁽¹²⁾ tests and Content Validity Index/CVI⁽⁶⁾.

Table 2. Cronbach coefficient, Spearman-Brown and CVI - Content Validity

Criteria	Cronbach's Alpha	Spearman-Brown	CVI
1 - Statements	1.00	1.00	1.00
2 - Ungrammatical Phrase	0.86	0.93	1.00
3 - Telegraphic Phrase	0.86	0.67	1.00
4 - Nominal Phrase	0.86	0.67	1.00
5 - Simple Period	0.86	0.93	1.00
6 - Compound Period	0.86	0.93	1.00
7 - Lack of Nominal Agreement	0.86	0.93	1.00
8 - Lack of Verbal Agreement	0.86	0.93	1.00
	0.84	0.83	1.00

DISCUSSION

Among the attributes most commonly used in the process of evaluating the psychometric properties of an instrument are validity and reliability. Validity can accurately measure the phenomenon to be studied and reliability allows to reproduce a result consistently even with different observers, representing how stable and consistent the instrument is⁽⁶⁾.

Table 1 shows the results of the application of the protocol by the judges using the statistical test of the Reliability Index. Authors argue that trustworthiness from 90% is appropriate⁽¹³⁾, however, some adopt an index from 70%. The formula used for calculations in this study says that every index equal to or greater than 70% is reliable⁽¹⁰⁾. Therefore, according to table 1, all reliability indexes were satisfactory.

In the criterion that deals with the number of statements (1), an index below 90% was obtained in the analysis of one of the five children, this occurrence directly influences the Reliability test, which, in these cases, tends to have a low percentage. Thus, if a judge finds two quantities of an item and the other only one, the Index is 50%, even if the quantities are very close⁽¹¹⁾. However, this does not imply considerably low values, since a minimum of twenty statements and a maximum of twenty-five were found, with a variation of only five. This question is much discussed and controversial since the Trust Index must not be observed only as a number, one must know about what is being collected and analyzed to then be able to interpret the presence or not of agreement between the judges^(10, 14).

As for the number of ungrammatical phrases (2), it is observed that this criterion reached a percentage below 90% in the analysis of three of the five children because the second judge found a lower number of phrases of this type.

Just like 1, the criterion that checks the number of noun phrases (4) reached an index below 90% in the analysis of one of the five children, because it had a minimum of 10 simple

periods and a maximum of 12 with a variation of only two periods, and this is also a reason for the decrease in the index.

Regarding the number of compound periods (6), a percentage below 90% was obtained in the analysis of only one of the five children, this also happened due to the low amount of deviation from the nominal agreement. Just like 6, the criterion that assesses the number of nominal agreement errors (7), a percentage below 90% was obtained in the analysis of two of the five children, due to the low amount of verbal agreement deviation found in children with Language Disorder. Criteria dealing with low values directly influence the test of the Reliability Index^(10,11). In summary, with the sum between the values, it was possible to apprehend that most of the criteria reached percentage values above 70% and this indicates that they are within the allowed variation of agreement between judges for reaching the minimum for accepting an item as pertinent.

The filing of the protocol criteria also allows for its improvement, in this sense, after the analyzes, but one criterion was added to the others, constituting the number 9: verification of the number of words in the statement. This criterion was added because it is pointed out as an important measure in the context of altered language development^(3,5), as children with language disorders produce sentences with fewer words⁽²⁾.

The statistical results described in table 2 show the acceptability, adequacy and relevance of each protocol criterion. There were conformity and compatibility in the responses of the specialists, which reveals that the MEP has reliability. The results of the statistical tests legitimize the reliability with indexes considered substantial both for the Alpha coefficient and for the Spearman. The CVI had a maximum concept of 1.0, this is because all the judges' responses were categorized according to the two largest Likert scales (partially agree and agree), which proves the effectiveness of the instrument's criteria.

Determining how rigorously the aspects of reliability and validity are addressed in a study is essential to guarantee the quality of an instrument, which helps the researcher to decide whether or not to apply the results in his clinical practice⁽¹⁵⁾.

Finally, it is worth remembering that reliability and validity are not fixed properties and, therefore, change according to the circumstances, population and purpose of the study. Measuring instruments unify clinical practice and research in different areas of knowledge, thus, assessing their quality is essential for the choice of instruments that provide valid and reliable measures.

In the clinical field, a systematic review examining the psychometric quality of various language assessment tools available to school-aged children emphasized the importance of the psychometric quality of procedures for speech-language therapists to make evidence-based decisions about the assessments they select when assessing children's language skills. In this sense, this study contributes to this premise, which is to use instruments with psychometric quality, such as validity and reliability.

It is worth mentioning that MEP will not be available until all of its validation steps are completed. It is being applied to a sample of typical children, whose results, in the future,

may serve as a parameter for the evaluation of children with a language disorder.

CONCLUSION

Regarding the content of the protocol, there was conformity and compatibility in the responses of the specialists, which reveals reliability. The results of the statistical tests legitimize the reliability with indexes considered substantial for both Alpha and Spearman coefficients. The CVI had a maximum concept of 1.0, which proves the effectiveness of the instrument's criteria.

REFERENCES

1. Hage SRV; Pinheiro LAC. Desenvolvimento típico de linguagem e sua importância para a identificação de suas alterações na infância. In: Lamônica DAC & Britto D. Tratado de linguagem: perspectivas contemporâneas. Ribeirão Preto: Booktoy. 2017. p.31-7.
2. Penke M. Syntax and language disorders. In: Tibor Kiss & Artemis Alexiadou (editors). Syntax – Theory and Analysis. An International Handbook. Handbook of Linguistics and Communication Science. Walter Gruyter Foundation: China. 2015. p. 1833-74.
3. Carvalho AMA, Befi-Lopes DM, Limongi SCO. Mean length utterance in Brazilian children: a comparative study between Down syndrome, specific language impairment, and typical language development. *CoDAS*. 2014; 26(3): 201-207. <http://dx.doi.org/10.1590/2317-1782/201420140516>
4. Santos ME, Lynce S, Carvalho S, Cacula M, Mineiro A. Mean length of utterance-words in children with typical language aged 4 to 5 years. *Revista CEFAC*. 2015; 17(4): 1143-1151. <http://dx.doi.org/10.1590/1982-021620151741315>
5. Denman D, Speyer R, Munro N, Pearce WM, Chen YW, Cordier R. Psychometric Properties of Language Assessments for children aged 4 - 12 years: a systematic review. *Frontiers in psychology*, 8: 1515, 2017. PMID: 28936189. DOI: 10.3389/fpsyg.2017.01515.
6. Alexandre NM, Coluci MZ. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. *Ciência & Saúde Coletiva*. 2011; 16(7): 3061-3068. <http://dx.doi.org/10.1590/S1413-81232011000800006>
7. Grant JS, Davis LL. Selection and use of content experts for instrument development. *Research in Nursing & Health*. 1997; 20(3): 269-274. PMID: 9179180. DOI: 10.1002/(sici)1098-240x(199706)20:3<269::aid-nur9>3.0.co;2-g.
8. Lynn, M R. Determination and quantification of content validity. *Nursing Research*. 1986; 35(6): 382-385. PMID: 3640358.
9. Likert, R. A Technique for the Measurement of Attitudes. *Archives of Psychology*. 1932; 140: 5-55.
10. Fagundes AJFM. Definição, descrição e registro do comportamento. 12ª ed. São Paulo: Edicon; 1999.
11. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951; 16(3): 297-334.
12. Spearman C. Correlation calculated from faulty data. *British Journal of Psychology*. 1910; 3(3): 271-295.
13. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. *Res Nurs Health* 2006; 29:489-497. PMID: 16977646. DOI: 10.1002/nur.20147.
14. Cunha CM, Almeida Neto OP, Stackfleth R. Principais métodos de avaliação psicométrica da confiabilidade de instrumentos de medida. *Revista de Atenção à Saúde*. 2016; 14(49):98-103. <https://doi.org/10.13037/ras.vol14n47.3391>
15. Souza Ac, Alexandre AMC, Guirardello EB. Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade. *Aplicações da Epidemiologia*. 2017; 26(3):649-659. <https://doi.org/10.5123/s1679-49742017000300022>.

Authors' contributions

LACP were responsible for collecting and analyzed the data, in addition to interpreting and elaborating the critical review of the study. PHC were

responsible for tabulating and interpreting the data statistically. SRVH were responsible for conceiving and outlining the study, as well as interpreting and developing the critical review of the work.