

Protocolo de Avaliação Morfossintática (PAM): validação de conteúdo

Morphosyntactic Evaluation Protocol (MEP): validation of content

Lorena Adami da Cruz Pinheiro¹ 

Anilton Pinheiro da Silva² 

Simone Rocha de Vasconcellos Hage³ 

Descritores

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

Keywords

Language Child
Language Development
Language Tests
Developmental Language Disorder
Validation Studies

Endereço para correspondência:

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

Recebido em: Julho 14, 2019.

Aceito em: Fevereiro 03, 2020.

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.

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.

Instituição onde o trabalho foi realizado: 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.

Conflito de interesses: nada a declarar.

Fontes de financiamento: FAPESP – Programas Regulares / Projeto de Pesquisa Regular – 2017/02990-9.



Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

INTRODUÇÃO

O conhecimento morfossintático é decisivo para que a criança evolua do nível de palavras isoladas para o de enunciados com uso de regras gramaticais da língua. Quando a criança passa a estabelecer relações entre as palavras numa frase e usar flexões por volta dos 18 meses, se instaura a morfossintaxe⁽¹⁾. No contexto clínico, este conhecimento é comumente afetado nos Transtornos da Linguagem, sejam eles primários ou não⁽²⁾. Apesar da importância que a morfossintaxe tem no desenvolvimento da comunicação infantil, ela tem sido negligenciada nos estudos brasileiros por não ser possível aferi-la de forma confiável no português do Brasil. No que tange à verificação especificamente da estrutura morfológica e sintática da linguagem, somente o MLU - *Mean Length Utterance* vem sendo empregado como procedimento de aferição nos últimos anos^(3,4). Neste contexto, a construção ou a adaptação transcultural de instrumentos de avaliação sintática representa um desafio para os avanços na avaliação da linguagem infantil. Para vencer este desafio é necessário que os instrumentos sejam validados para falantes brasileiros.

No intuito de contribuir com a avaliação morfossintática de crianças cujo idioma é o português do Brasil, protocolo de avaliação morfossintática (PAM) foi desenvolvido por pesquisadores brasileiros e envolveu duas etapas: primeiramente, foram elencadas categorias presentes nos modelos evolutivos de aquisição morfossintática com base em revisão bibliográfica. A revisão buscou em diferentes fontes bibliográficas informações sobre avaliação morfossintática de crianças e sobre a gramática da língua portuguesa. Foram consideradas as regras implícitas do uso da língua e as fases de aquisição da linguagem de crianças entre 2;6 e 5;0, faixa etária que o protocolo abrange. Numa segunda etapa, o protocolo foi aplicado em um grupo piloto constituído tanto por crianças com desenvolvimento típico de linguagem, como com Transtorno de Linguagem primário, ou seja, não associado a uma condição biomédica. Esta aplicação foi fundamental para verificar se as categorias identificadas na literatura eram compatíveis com o corpus da fala de crianças com transtorno. Com base na estrutura gramatical da língua portuguesa e nas características mais ocorrentes da construção morfossintática de crianças com alteração de linguagem, o PAM elegeu inicialmente oito critérios de análise para serem verificados sobre a transcrição ortográfica de conversação entre criança e interlocutor adulto por 20 minutos. Os critérios de análise foram: 1. Quantidade de enunciados produzidos, 2. Quantidade de frases agramaticais, 3. Quantidade de frases telegráficas, 4. Quantidade de frases nominais, 5. Quantidade de períodos simples, 6. Quantidade de períodos compostos, 7. Quantidade de erros de concordância nominal e 8. Quantidade de erros de concordância verbal.

Entretanto, para que um instrumento de avaliação criado possa ser utilizado com confiabilidade é fundamental que se verifique suas qualidades psicométricas. Cada vez mais se enfatiza a importância da verificação destas qualidades nos procedimentos de avaliação de linguagem⁽⁵⁾, e uma destas qualidades é a validade em que se verifica se um instrumento

mede precisamente o que se propõe a medir⁽⁶⁾. Existem várias formas de validar um instrumento, uma delas é a de conteúdo. A validade de conteúdo refere-se ao grau em que o conteúdo de um instrumento reflete o construto que está sendo medido e implica em avaliação com quantificação de julgamentos, tendo como principal objetivo o de avaliar e aperfeiçoar os critérios utilizados num instrumento⁽⁷⁾.

O objetivo deste estudo foi verificar a validade de conteúdo de um instrumento criado por brasileiros para verificação das habilidades morfossintáticas de crianças em desenvolvimento de linguagem.

MÉTODO

A pesquisa teve a aprovação do Comitê de Ética em Pesquisa em Seres Humanos da Faculdade de Odontologia de Bauru (CAAE 68562317.4.0000.5417). Os juizes assinaram o Termo de Consentimento Livre e Esclarecido (TCLE), conforme Resolução 466/2012.

Os estudos sobre validação de conteúdo são divergentes no que diz respeito a quantidade de avaliadores. A recomendação pode variar de dois a vinte⁽⁷⁾. Lynn⁽⁸⁾ aponta que a quantidade mínima de juizes pode ser três. Assim, a escolha foi por três avaliadores especialistas com experiência na avaliação de todos os aspectos abordados no protocolo. Isto posto, convidamos três especialistas no campo da linguagem, dois fonoaudiólogos e um linguista que se dispuseram a aplicar em cinco amostras de fala e avaliar o instrumento.

A validação de conteúdo foi realizada em duas etapas. Primeiramente os juizes receberam a transcrição da amostra de fala de cinco crianças com Transtorno de Linguagem, por ser a fala delas o objetivo final da criação do protocolo, e aplicaram o protocolo com base nas orientações fornecidas por meio de um manual “passo a passo”. Após a aplicação, os juizes avaliaram o instrumento por meio de um questionário com perguntas referentes à pertinência, relevância, aplicabilidade e representatividade de cada critério do protocolo. As perguntas sobre cada item serviram para auxiliar no julgamento e na própria avaliação dos juizes. O uso de um questionário para obtenção do julgamento dos especialistas é um método que minimiza os vieses e padroniza as informações solicitadas sobre o conteúdo de cada item⁽⁷⁾.

Para a análise estatística do questionário foi usada a escala tipo *Likert*⁽⁹⁾ com quatro níveis de suporte: A – Discordo; B – Discordo parcialmente; C – Concordo parcialmente; D – Concordo.

Com os dados da aplicação do protocolo pelos juizes foi gerada tabela a qual aplicou-se o cálculo de concordância entre eles ou o índice de fidedignidade. O cálculo foi realizado utilizando-se a técnica em que o índice de confiabilidade deve ser igual ou maior que 70%⁽¹⁰⁾, o que recomenda um fato não ser produzido ao acaso, portanto indicando fidedignidade.

No rol de respostas do questionário de cada juiz foi aplicado o teste de *Cronbach*⁽¹¹⁾, medida estatística que assegura a consistência interna de um teste ou escala. *Cronbach*⁽¹¹⁾ quantificou essa confiabilidade propondo um coeficiente, α , que varia de 0 a 1. Se α está próximo de 0, então as respostas qualificadas não são

confiáveis, e se estiver próximo de 1, as respostas são muito confiáveis. Se $\alpha \geq 0,8$, então as respostas são consideradas confiáveis.

A confiabilidade do instrumento também foi verificada por meio do método das metades, ou seja, os itens que compõem o PAM foram agrupados em duas metades, comparando-se as pontuações obtidas a estas metades. Foram estimados coeficientes de precisão por meio da fórmula de *Spearman-Brown*⁽¹²⁾.

RESULTADOS

A Tabela 1 representa os dados obtidos por meio da aplicação do protocolo pelos juízes numa amostra de fala já transcrita de cinco crianças com Transtorno de Linguagem. A análise das respostas ocorreu sobre oito critérios do protocolo.

Tabela 1. Índice de Fidedignidade entre os juízes sobre a aplicação do protocolo numa amostra de fala espontânea de cinco crianças

Critérios do PAM	Índice de Fidedignidade (IF) entre juízes sobre a aplicação do protocolo amostras de fala espontânea														
	Índice Fidedignidade					Índice Fidedignidade					Índice Fidedignidade				
	Juízes 1x2					Juízes 1x3					Juízes 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

Legenda. C = criança. 1. Quantidade de enunciados produzidos, 2. Quantidade de frases agramaticais, 3. Quantidade de frases telegráficas, 4. Quantidade de frases nominais, 5. Quantidade de períodos simples, 6. Quantidade de períodos compostos, 7. Quantidade de erros de concordância nominal e 8. Quantidade de erros de concordância verbal.

A Tabela 2 é referente aos resultados obtidos com a aplicação dos testes Alpha de *Cronbach*⁽¹¹⁾, *Spearman-Brown*⁽¹²⁾ e Índice de Validade de Conteúdo/IVC⁽⁶⁾.

Tabela 2. Coeficiente de *Chonbach*, *Spearman-Brown* e IVC – Validade de Conteúdo

Critério	Alpha de Cronbach	Spearman-Brown	IVC
1 - Enunciado	1,00	1,00	1,00
2 - Frase Agramatical	0,86	0,93	1,00
3 - Frase Telegráfica	0,86	0,67	1,00
4 - Frase Nominal	0,86	0,67	1,00
5 - Período Simples	0,86	0,93	1,00
6 - Período Composto	0,86	0,93	1,00
7 - Falta de Concordância Nominal	0,86	0,93	1,00
8 - Falta de Concordância Verbal	0,86	0,93	1,00
	0,84	0,83	1,00

DISCUSSÃO

Entre os atributos mais comumente utilizados no processo de avaliação das propriedades psicométricas de um instrumento estão a validade e a confiabilidade. A validade é capaz de medir com precisão o fenômeno a ser estudado e a confiabilidade permite reproduzir um resultado de forma consistente mesmo com observadores diferentes, representando o quanto o instrumento é estável e consistente⁽⁶⁾.

A tabelas 1 apresenta os resultados da aplicação do protocolo pelos juízes com o emprego do teste estatístico Índice de Fidedignidade. Autores defendem ser apropriado a fidedignidade a partir de 90%⁽¹³⁾, entretanto, há os que adotam índice a partir de 70%. A fórmula utilizada para os cálculos deste estudo diz ser fidedigno todo índice igual ou maior que 70%⁽¹⁰⁾. Portanto, de acordo com a tabela 1, todos os índices de fidedignidade foram satisfatórios.

No critério que trata da quantidade de enunciados (1) obteve-se índice abaixo de 90% na análise de uma das cinco crianças, essa ocorrência influencia diretamente no teste de Fidedignidade, a qual, nesses casos, tende a ter baixa

porcentagem. Assim, se um juiz encontra duas quantidades de um item e o outro apenas uma, o Índice é de 50%, mesmo se tratando de quantidades muito próximas⁽¹¹⁾. Porém, isto não implica em valores consideravelmente baixos, posto que foi encontrado valor mínimo de vinte enunciados e máximo de vinte e cinco com variação de apenas cinco. Esta questão é muito discutida e controversa, pois o Índice de Fidedignidade não deve ser observado apenas como um número, deve-se ter um conhecimento do que está sendo coletado e analisado para, então, poder interpretar a presença ou não de concordância entre os juízes^(10, 14).

Quanto a quantidade de frases agramaticais (2), observa-se que este critério atingiu porcentagem abaixo de 90% na análise de três das cinco crianças pelo fato do segundo juiz ter encontrado menor número de frases deste tipo.

Assim como o 1, o critério que verifica a quantidade de frases nominais (4) atingiu índice abaixo de 90% na análise de uma das cinco crianças, isso porque teve o mínimo de 10 período simples e máximo de 12 com variação de apenas dois períodos e esse também é um motivo para diminuição do índice.

Em relação a quantidade de períodos compostos (6), obteve-se porcentagem abaixo de 90% na análise de apenas uma das cinco crianças, isso também aconteceu em virtude da baixa quantidade de desvio de concordância nominal. Assim como o 6, o critério que avalia a quantidade de erros de concordância nominal (7), obteve-se porcentagem abaixo de 90% na análise de duas das cinco crianças, decorrente da baixa quantidade de desvio de concordância verbal encontrado nas crianças com Transtorno de Linguagem. Critérios que tratam de baixos valores influenciam diretamente no teste do Índice de Fidedignidade^(10,11). Em resumo, com a soma entre os valores foi possível apreender que a maior parte dos critérios atingiu valores percentuais acima de 70% e isso indica que estão dentro da variação permitida de concordância entre juízes por atingir o mínimo para aceitação de um item como pertinente.

O ajuizamento dos critérios do protocolo também permite seu aprimoramento, neste sentido, após as análises, mas um critério foi somado aos outros, constituindo o de número 9: verificação da quantidade de palavras no enunciado. Este critério foi acrescentado por ser apontado como uma medida importante no contexto da alteração do desenvolvimento de linguagem^(3,5), na medida que crianças com Transtornos de linguagem produzem frases com menor número de palavras⁽²⁾.

Os resultados estatísticos descritos na tabela 2 mostram a aceitabilidade, adequação e relevância de cada critério do protocolo. Houve conformidade e compatibilidade nas respostas dos especialistas o que revela que o PAM tem confiabilidade. Os resultados dos testes estatísticos legitimam a fidedignidade com índices considerados substanciais tanto para coeficiente *Alpha*, quanto para o Spearman. O IVC teve conceito máximo de 1,0, isso pela razão de todas as respostas dos juízes estarem categorizadas de acordo com as duas maiores escalas de *Likert* (concordo parcialmente e concordo) o que comprova a eficácia dos critérios do instrumento.

Determinar quão rigorosamente os aspectos de confiabilidade e validade são abordados em um estudo é essencial para garantir a qualidade de um instrumento, o que auxilia o pesquisador a decidir se deve ou não aplicar os resultados em sua prática clínica⁽¹⁵⁾.

Por fim, vale lembrar que a confiabilidade e a validade não são propriedades fixas e, portanto, alteram de acordo com as circunstâncias, população e finalidade do estudo. Os instrumentos de medida unificam a prática clínica e a pesquisa em diferentes áreas do conhecimento, assim, a avaliação de sua qualidade é essencial para a eleição de instrumentos que forneçam medidas válidas e confiáveis.

No campo clínico, revisão sistemática em que se examinou a qualidade psicométrica de instrumentos diversos de avaliações de linguagem disponíveis para crianças em idade escolar enfatizou a importância da qualidade psicométrica dos procedimentos para que os fonoaudiólogos tomem decisões baseadas em evidências sobre as avaliações que eles selecionam ao aferir as habilidades de linguagem das crianças. Neste sentido, este estudo vem contribuir com esta premissa, que é a de se utilizar instrumentos com qualidade psicométrica, como a validade e a confiabilidade.

Vale ressaltar que o PAM não estará disponível até que todas as suas etapas de validação estejam concluídas. Ele está sendo aplicado em amostra de crianças típicas, cujos resultados, no futuro, poderão servir de parâmetro para a avaliação de crianças com transtorno de linguagem.

CONCLUSÃO

Em relação ao conteúdo do protocolo, houve conformidade e compatibilidade nas respostas dos especialistas o que revela confiabilidade. Os resultados dos testes estatísticos legitimam a fidedignidade com índices considerados substanciais tanto para coeficiente *Alpha* e Spearman. O IVC teve conceito máximo de 1,0, o que comprova a eficácia dos critérios do instrumento.

REFERÊNCIAS

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>.

Contribuição dos autores

LACP coletou e analisou os dados, além de ter interpretado e elaborado a revisão crítica do estudo. APS tabulou e interpretou os dados estatisticamente. SRVH concebeu e delimitou o estudo, assim como interpretou e elaborou a revisão crítica do trabalho.