Brief Communication

Comunicação Breve

Recommendations for elaboration, transcultural adaptation and validation process of tests in Speech, Hearing and Language Pathology

Recomendações para elaboração, tradução, adaptação transcultural e processo de validação de testes em Fonoaudiologia

Leandro Pernambuco[1]
Albert Espelt[2,3,4]
Hipólito Virgílio Magalhães Junior[5]
Kenio Costa de Lima[5]

Keywords

Validation Studies
Validity of Tests
Speech, Language and Hearing Sciences
Psychometrics
Evaluation

ABSTRACT

Objective: to present a guide with recommendations for translation, adaptation, elaboration and process of validation of tests in Speech and Language Pathology. **Methods:** the recommendations were based on international guidelines with a focus on the elaboration, translation, cross-cultural adaptation and validation process of tests. **Results:** the recommendations were grouped into two Charts, one of them with procedures for translation and transcultural adaptation and the other for obtaining evidence of validity, reliability and measures of accuracy of the tests. **Conclusion:** a guide with norms for the organization and systematization of the process of elaboration, translation, cross-cultural adaptation and validation process of tests in Speech and Language Pathology was created.

Descritores

Estudos de Validação
Validade dos Testes
Fonoaudiologia
Psicometria
Avaliação

RESUMO

Objetivo: apresentar um guia com recomendações para a tradução, adaptação, elaboração e processo de validação de testes em Fonoaudiologia. **Método:** as recomendações apresentadas foram baseadas em diretrizes internacionais tradicionais cujo enfoque está na elaboração, tradução, adaptação transcultural e processo de validação de testes. **Resultado:** as recomendações foram compiladas em dois quadros, sendo um deles referente aos procedimentos para tradução e adaptação transcultural e o outro à obtenção de evidências de validade, confiabilidade e medidas de acurácia dos testes. **Conclusão:** foi apresentado um guia com as principais recomendações para a organização e sistematização do processo de elaboração, tradução, adaptação transcultural e processo de validação de testes em Fonoaudiologia.

Correspondence address:
Leandro Pernambuco
Departamento de Fonoaudiologia, Universidade Federal da Paraíba – UFPBCidade Universitária, Bairro Castelo Branco, João Pessoa (PB), Brazil,
CEP: 58051-900.
E-mail: leandroape@globo.com

Received: October 25, 2016

Accepted: November 07, 2016

Study carried out at Departamento de Fonoaudiologia, Universidade Federal da Paraíba – UFPB - João Pessoa (PB), Brazil.

[1] Universidade Federal da Paraíba – UFPB - João Pessoa (PB), Brazil.

[2] Agència de Salut Pública de Barcelona – ASPB - Barcelona, Spain.

[3] Universitat Autònoma de Barcelona – UAB - Bellaterra, Spain.

[4] Consorcio de Investigación Biomédica en Red de Epidemiologia y Salud Pública – CIBERESP.

[5] Universidade Federal do Rio Grande do Norte – UFRN - Natal (RN), Brazil.

Financial support: nothing to declare.

Conflict of interests: nothing to declare.

INTRODUCTION

In the area of health care, an adequate evaluation process depends on the utilization of tests whose interpretations of the results are valid, reliable/precise and equitable. Validity refers to the gathering of evidence, which indicates if the test really measures what it intends to measure. Reliability/precision indicates if the test is reproducible over time (stability), if there is control of measurement errors (precision) and if the result of the test is dependent on the items that constitute it (homogeneity)[1-3]. Equity allows the analysis of whether the test evaluates the individuals in an impartial manner, without permitting irrelevant issues from unduly exercising influence on the outcome and generating inequalities[1,4].

In Brazil, the application of these principals by Speech-language Therapy is rare or undertaken in a partial manner, with a systematization of the methodological procedure still being necessary[5]. The validation process is not satisfied simply with the elaboration, translation and adaptation of a test. For this to occur, the utilization of international norms, which guarantee the effective obtainment of the psychometric and clinimetric properties of the test, are necessary.

Despite there being controversies between some guidelines[6], certain methodological principals cannot be neglected. The aim of this manuscript is to present a short guide with recommendations for the translation, adaptation, elaboration and validation process for Speech-language therapy tests.

METHODS

The recommendations presented here were based on the experience of the authors with the most traditional and commonly utilized international guidelines related to the elaboration, translation, transcultural adaptation and validation process for tests. In the case of translation and transcultural adaptation, the recommendations of Beaton et al.[7] were followed, as well as the guidelines of the *International Test Commission* (ITC)[8]. In the case of the validation process for tests, the principals of the *Standards for Education and Psychological Testing* (SEPT)[1], were used, a guideline proposed by three North-American organizations, which brings together the most reliable and utilized recommendations and definitions related to the psychometric aspects involved, from the elaboration to interpretation of the tests.

Given that this manuscript is a methodological note, which did not involve collection of data from human beings, it was not necessary to submit the study for appraisal by an ethics committee.

RESULTS

Recommendations for the elaboration, translation, transcultural adaptation and validation process for Speech-language therapy tests are presented in Charts 1 and 2.

Chart 1. Procedures for translation and transcultural adaptation of tests

Step	Procedures
a) Initial Guidelines	***a1)*** Obtainment of permission from the authors of the original test. ***a2)*** Formation of a committee of specialists (authors of the new version and other specialists + at least one author of the original version, if possible) to debate the concepts associated with the test to be adapted, considering the characteristics of the population and target culture. Participants should be aware of the test objectives.
b) Development of the test	***b1) Translation:*** Two trained translators, native speakers of the target language and fluent in the source culture and language, independently translate the test, taking into account conceptual equivalence and avoiding literal translation. It is recommended that both translators do not know the test and that they are not specialists in the outcome investigated by the test so that the representativity of the common usage of the target language is maintained. ***b2) Synthesis of the translations:*** undertaken in a collaborative manner*, preferably by the same committee mentioned in procedure a2. The committee will construct a single version from the comparison of the translations and evaluation of semantic, idiomatic, conceptual, linguistic and contextual discrepancies. ***b3) Applicability of the synthesis of translations/operational equivalence:*** the adequacy, structure and application of the items should be verified in a real-life context. There is no explicit recommendation for the sample size at this stage. The formation of representative strata of the target population, with at least ten individuals in each strata, is suggested. A cognitive interview is performed to verify if the population understands the test items. If the test is a questionnaire, the paraphrasing strategy is recommended, in which the interviewer asks the questions and requests the interviewee to immediately repeat them back. The interviewer should also observe possible operational difficulties (for example: long application time; excessive number of questions), non-verbal reactions of the interviewee (for example: boredom; facial expressions) and suggestions from the interviewee to improve the understanding of the item. The committee mentioned in procedure b2 will meet once again if changes to the test are necessary. The test will be reapplied with new strata and this procedure will be repeated until the test is applied without any difficulty and with comprehension. ***b4) Back-translation or reverse translation:*** the version obtained in procedure b3 will be translated into the source language to evaluate if the items reflect the content of the original version. The version in the target language will be sent to at least two trained translators who do not know the test and who are native speakers of the source language and fluent in the target language and culture. ***b5) Synthesis of the back-translated versions***: undertaken in a collaborative manner*, preferably by the same committee mentioned in procedure a2. The committee will compare the retranslations to the original test and evaluate the semantic, idiomatic, conceptual, linguistic and contextual discrepancies. *During this step, it is recommended that the authors of the original version give their opinion in terms of the retranslated version.* ***b6) Final synthesis:*** undertaken in a collaborative manner*, preferably by the same committee mentioned in procedure a2. The committee will compare the original and final versions in the target language in terms of semantic, idiomatic, conceptual, linguistic, experiential and contextual equivalences.

*In the collaborative evaluations, assessment strategies can be adopted in relation to the equivalencies utilizing, for example, a visual analogical scale or scales of the Likert type. The calculation of the Content Validity Index for Scales (CVI-S) and for Item (CVI-I)[9] helps at this stage as a quantitative indicator of agreement between the evaluators

Chart 2. Procedures for the obtainment of evidence of validity, reliability and measures of accuracy of the tests

Step	Procedure
Evidence of validity based on test content	Content refers to the themes, writing, formatting, tasks or questions in the test, as well as instructions for the procedures necessary to administer and mark it[1,3]. Some strategies to elaborate tests are: extensive literature review; empirical experience of the researchers with the outcome of interest; interviews with key informants; holding a panel with specialists and the target population; and consultation with the target population using focus groups. In the elaboration of the items, syntactic and semantic aspects, which contribute to the clarity, pertinence, coherence and scope of the items, as well as operational aspects should be considered. The representativity and relevance of the items in terms of the outcome should be evaluated by a committee of at least ten evaluators with expertise in the theme of the test[2,3]. The use by the evaluators of a visual analogical scale or scales of the Likert type to evaluate the test items is recommended. The Content Validity Index for Scales (CVI-S) and for Item (CVI-I)[9] for the presentation of the quantitative indicator of agreement between the evaluators is calculated.
Evidence of validity based on the response processes	The response processes result from the observations or judgments regarding the behavior or performance of different strata of the target population during the application of the test. During this step, we seek to understand which psychological, cognitive and social processes are involved in the application of the test. The procedures are similar to those described in b3 in Chart 1.
Evidence of validity based on internal consistency	The extent of the relation between the test items and the outcome, judged from the application of the test in a sample of the target population, is verified. If the internal dimensions of the test are theoretically defined by the authors, a confirmatory factorial analysis (CFA) should be carried out. If the authors need to investigate the possibility of cutting items or the existence of different dimensions of the test, principal component analysis (PCA) or factorial exploratory analysis (FEA), with the posterior CFA to validate the model found, should be undertaken. If the researcher follows the Item Response Theory (IRT) to investigate the latent traces, they can perform this analysis after the CFA. Each one of the analyses mentioned requires a different sample. Other parameters generally adopted are the corrected item-total correlation and the inter-item correlation with values above 0.3[2]. Another characteristic investigated during this phase is the Differential Functioning of Items (DFI), which verifies if variables, either individually or in group, place one or more categories at an advantage or disadvantage in relation to the other/s[4].
Evidence of validity based on the relation to other variables	The extent to which the relations between the instrument and the other external variables are consistent with the outcome is verified. 1) *Convergent Validity:* analysis of the relations between the test result and other measures which evaluate the same or similar outcomes; 2) *Discriminant Validity*: the test is evaluated in terms of its capacity to differentiate distinct groups from the target population or compare it to other measures with different outcomes; 3) *Criterion Validity*: the criteria are an operationally distinct attribute or result of the outcome of the test and are established by way of hypotheses formulated by the researchers. In the *concurrent validity*, the result of the test is compared to the outcomes of other tests administered in parallel and, in the *predictive validity*, the test result in relation to the condition of the evaluated individual in the future or the result of subsequent evaluations is evaluated.
Reliability/ Precision	The test should be administered at two distinct times, generally between 7 to 14 days. This period is variable according to the characteristics of the outcome, especially its stability over time. The test should be applied by the same researcher and distinct researchers at both times. The Intraclass Coefficient Correlation (ICC) or Kappa index to verify inter-evaluator and intra-evaluator agreement will be calculated. Internal consistency is obtained using the values of Cronbach's alpha, measuring error, item-total correlation and inter-item correlation.
Test equity	Specialists should assess whether the result previously obtained in the DFI analysis is reproduced in a just manner and guarantee that the results in the test are not influenced by possible disadvantages caused by individual variables.
Accuracy	To obtain the indicators for sensitivity, specificity, positive and negative predictive value and positive and negative likelihood ratio, one test result will be compared to the other, taking into account the standard of reference. The cutoff point of the instrument and its discriminatory power in discriminating people with and without the outcome will be obtained using the ROC curve (receiver operator characteristic curve).
Evidence of validity based on test consequences	Characterized by the accumulation of evidence related to the use and efficacy of the test. Includes analyses of the responsiveness of the test to certain interventions and takes into account subsequent studies for application and reproduction of the test in practical contexts and related to the outcome.

DISCUSSION

The elaboration, translation, transcultural adaptation and validation process are methodological procedures necessary to guarantee that the interpretation of the test results is valid and reliable. In Brazil, these procedures have been frequently reproduced in Speech-language therapy, however not always with the indispensable scientific rigor recommended by the international guidelines.

In the case of translation and transcultural adaptation, there are around thirty guidelines, but no consensus pointing toward a single standard of reference[6]. Recommendations proposed in the 1990s and updated in the 2000s are the most disseminated in the international literature[7]. In the present study these recommendations were followed, as well as the guidelines from the *International Test Commission Guidelines* (ITC)[8].

The SEPT[10] exists since the 1950s and specifies, beyond reliability/precision and equity, five sources of evidence for validity, based on: content, response processes, internal structure, relation with other variables and consequences of the test. In the area of health care, the proposals of the *Scientific Advisory Committee of the Medical Outcomes Trust* (SAC)[10] and of

the *Consensus-based Standards for the Selection of Health Measurement Instruments* (COSMIN)[11] are referenced with frequency and can also be consulted. However, it is worth noting that these two guidelines consider psychometric concepts that the SEPT already ceased to advocate since the 1999 edition, that is to say, they are proposals, which are should be analyzed with parsimony, especially due to their conceptual mark.

CONCLUSION

The main recommendations for the organization and systematization of the processes of elaboration, translation, transcultural adaptation and validation for Speech-language therapy tests are presented. We recommend a careful reading for each one of the procedures mentioned here, but we hope that these guidelines aid in the realization of future research in the area and can help speech-language therapists choose their tests in a more exacting fashion, assembling the best evidence to evaluate a determined outcome.

REFERENCES

1. AERA: American Educational Research Association, APA: American Psychological Association, NCME: National Council on Measurement in Education. Standards for educational and psychological testing. New York: AERA; 2014.
2. Streiner DL, Norman GL. Health measurement scales: a practical guide to their development and use. 4th ed. New York: Oxford University Press; 2008.
3. Abad FJ, Olea J, Ponsonda V, Garcia C. Measurement in social sciences and health. Madrid: Sintesis; 2011.
4. Espelt A, Viladrich C, Doval E, Aliaga J, García-Rueda R, Tárrega S. Uso equitativo de tests en ciencias de la salud. Gac Sanit. 2014;28(5):408-10. PMid:24928357. http://dx.doi.org/10.1016/j.gaceta.2014.05.001.
5. Gurgel LG, Kaiser V, Reppold TZ. A busca de evidências de validade no desenvolvimento de instrumentos em Fonoaudiologia: revisão sistemática. Audiol Commun Res. 2015;20(4):371-83. http://dx.doi.org/10.1590/2317-6431-2015-1600.
6. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. J Clin Epidemiol. 2015;68(4):435-41. PMid:25698408. http://dx.doi.org/10.1016/j.jclinepi.2014.11.021.
7. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000;25(24):3186-91. PMid:11124735. http://dx.doi.org/10.1097/00007632-200012150-00014.
8. Muñiz J, Elosua P, Hambleton RK. Directrices para la traducción y adaptación de los tests: segunda edición. Psicothema. 2013;25(2):151-7. PMid:23628527.
9. Polit DF, Beck CT. The content validity index: are you sure you know what's being reprted? Critique and recommendations. Res Nurs Health. 2006;29(5):489-97. PMid:16977646. http://dx.doi.org/10.1002/nur.20147.
10. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality of life instruments: attributes and review criteria. Qual Life Res. 2002;11(3):193-205. PMid:12074258. http://dx.doi.org/10.1023/A:1015291021312.
11. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45. PMid:20494804. http://dx.doi.org/10.1016/j.jclinepi.2010.02.006.

Author contributions

LP and HVMJ contributed with the conception, writing and final revision of the article; AE and KCL contributed with the final revision of the article.