

Validation of the Objective Structured Assessment of Technical Skill in Brasil

 Marcelo Esteves Chaves Campos¹
 Marcelo Magaldi Ribeiro de Oliveira¹
 Lilian Bambirra de Assis²
 Augusto Barbosa Reis¹
 Flávio Bambirra Gonçalves¹

1. Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

2. Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, MG, Brasil

<http://dx.doi.org/10.1590/1806-9282.66.3.328>

SUMMARY

BACKGROUND: The aim of this study was to perform a cross-cultural adaptation of the Objective Structured Assessment of Technical Skill (OSATS) tool into Brazilian Portuguese and to determine its reproducibility and validity in Brasil.

METHODS: A Brazilian Portuguese version of OSATS was created through a process of translation, back-translation, expert panel evaluation, pilot testing, and then its validation. For the construct and the concurrent validities, twelve participants were divided into a group of six experts and six novices, who had to perform tasks on a simulation model using human placentas. Each participant was filmed, and two blinded raters would then evaluate their performance using the traditional subjective method and then the Brazilian Portuguese version of OSATS.

RESULTS: The Brazilian Portuguese version of OSATS had the face, content, construct, and concurrent validities achieved. The average experts' score and standard deviations were 34 and 0.894, respectively, for Judge 1 and 34.33 and 0.816 for Judge 2. In the case of novices, it was 13.33 and 2.388 for Judge 1 and 13.33 and 3.204 for Judge 2. The concordance between the judges was evident, with the Correlation Coefficient (Pearson) of 0.9944 with CI 95% between 0.9797 and 0.9985, with $p < 10^{-10}$, evidencing the excellent reproducibility of the instrument.

CONCLUSION: This preliminary study suggests that the Brazilian Portuguese version of OSATS can reliably and validly assess surgical skills in Brasil.

KEYWORDS: Educational measurement. Simulation. Training. Surgical procedures, operative/education. Surveys and questionnaires.

INTRODUCTION

Traditionally, the proficiency of surgical residents is based on case records that measure their operational experience but do not assess their performance^{1,2}. In addition, the performance assessment of residents is often carried out based on subjective criteria, which have low reliability and hinder the feedback to the learner³.

In recent years, there has been a shift in paradigm

regarding surgical education and the profile of the residents⁴⁻⁶. A systematic procedure has been increasingly used to assess operative performance objectively^{7,8}. For this purpose, the instrument most used worldwide is the Objective Structured Assessment of Technical Skills (OSATS)^{9,10}. It is composed of a global rating scale (GRS-OSATS) with seven assessment items scored on a Likert scale of 5 points. Therefore, the total score of

DATE OF SUBMISSION: 02-Sep-2019

DATE OF ACCEPTANCE: 10-Oct-2019

CORRESPONDING AUTHOR: Marcelo Esteves Chaves Campos

Rua Congonhas 792 - 301 Santo Antônio - Belo Horizonte, Minas Gerais, Brasil, 30330-100

Tel.: 55 31 991967784

E-mail: camposmec@yahoo.com.br

the GRS-OSATS varies from 7 to 35, with higher scores indicating a greater technical ability of the surgeon¹¹.

The GRS-OSATS was originally developed in English at the University of Toronto, and surgical researchers who wish to use this tool in a country in a language other than English have to carry out a process of transcultural adaptation and validation¹²⁻¹⁴. As far as we know, no validation study for the scale has been published in Brasil yet. Thus, the objective of this work was to cross-culturally adapt the GRS-OSATS for Brazilian Portuguese and validate it in Brasil.

METHODS

This cross-sectional observational study was approved by the Research Ethics Committee (Coep - CAAE Decision No: 0364.0.203.000-11), Federal University of Minas Gerais (UFMG), Brasil. All participants consented in writing to participating in this work.

We collected 12 human placentas from the obstetrics department of the Hospital das Clínicas of the Federal University of Minas Gerais (HC-UFMG). The pregnant women were submitted to a pre-natal infection assessment and signed the consent for the placenta donation for the practice of surgical techniques. The placentas were returned in their entirety to the Pathology Department of the UFMG five days after they were obtained, and their partial or total use for other purposes was forbidden.

The research was divided into two stages: first, the GRS-OSATS was transculturally adapted for Brazilian Portuguese; then, a validation study was conducted to determine its reproducibility, validity, and reliability in Brasil.

Transcultural adaptation

The cross-cultural adaptation of the original version of the GRS-OSATS was divided into five stages: the initial translation, translation synthesis, back-translation, consensual version, and pre-test¹⁵.

The first stage was the initial translation of the original version of the GRS-OSATS from English into Brazilian Portuguese by two bilingual translators. These produced two different versions of the translation, T1 and T2.

The second stage was the translation synthesis. The T1 and T2 versions were discussed with both translators, thereby producing a single version in Brazilian Portuguese, i.e., version T1-2.

In the third stage, two bilingual translators, who had no previous contact with the original instrument, back-translated separately the T1-2 version into English, which resulted in versions BT1 and BT2. These new versions in English allowed the identification of possible translation errors and grammar inconsistencies in comparison with the original version.

The fourth stage consisted of the assessment of all versions by a committee composed by the authors of this study. The objective of this stage was to review all translations to get a final single version in Brazilian Portuguese (FV-GRS-OSATS). Thus, we evaluated the semantic equivalence with respect to the meanings of words with attention to idiomatic expressions and colloquialisms; the experimental equivalence, comparing the realities of different countries and cultures; and the conceptual equivalence, ensuring that words have the same definition.

The last stage was the pre-test for adjustments and detection of inconsistencies and to allow the validation of the instrument. In the pre-testing, the FV-GRS-OSATS was presented to 20 surgeons. After examining it, the surgeons were asked about the difficulty in interpreting or understanding the instrument. They were then requested to assess the clarity of the items using a Likert scale of 5 points (1, not clear at all; 2, not very clear; 3, somewhat clear; 4, clear; and 5, very clear).

Validation

After the pre-test, we started the validation stage, in which we handed out to 20 participants a questionnaire on the ability of an instrument to measure the residents' technical skills in a general way for face validity and specifically for content validation of each item. A Likert scale of 5 points was also used (1, not capable; 2, barely capable; 3, moderately capable; 4, capable; and 5, very capable). The reliability, i.e., the consistency between evaluators, was measured by calculating Cronbach's coefficient alpha, whereas a value above 0.70 was acceptable.

For the construct validation (ability to differentiate the performance between experts and beginners) and the concurrent validation (comparison of the traditional subjective method with the proposed one), we recruited 12 participants, divided into two groups: six experts in surgery for the Expert Group (EG), which comprised surgeons with over ten years of experience, and six beginners for the Novice Group (NG), which comprised 1st-year surgical residents with little or no surgical experience.

As seen in Figure 1, a total of 12 human placentas were prepared, according to Oliveira Magaldi et al.¹⁶. Then, a standardized explanation of basic surgical techniques exercises in these models of training was presented to both groups. These consisted of dissecting a blood vessel, sectioning it with surgical scissors, and performing a hemostatic suture. Thus, we began the practice rounds using surgical simulation models.

Each training session was filmed, with special attention so that the camera framing captured only the hands of the participant during the exercises. The videos were then seen separately watched at two distinct moments by two experts in surgical education chosen as judges and who were not present during the training sessions. They were unaware of the participants' experience levels. Initially, the two judges evaluated the videos in the traditional subjective way and graded them from A to D

(A, excellent; B, good; C, regular; D, bad. After 15 days, both evaluated the participants' videos again; however, this time, using the FV-GRS-OSATS. The grades were distributed as follows: A for scores from 28 to 35, B for scores from 21 to 27, C for scores from 14 to 20, and D for scores from 7 to 13. The Correlation Coefficient (Pearson) between both judges was calculated with the help of Stata version 11.0, and we considered a CI of 95% and p<0.05.

RESULTS

In the first stage of the GRS-OSATS transcultural adaptation, there was a difference in the translation of certain words. However, since the meanings were preserved, the translation synthesis and back-translation stages were carried out without questions. In

TABLE 1. FINAL VERSION OF THE GRS-OSATS IN BRAZILIAN PORTUGUESE

Escala de Classificação Global de Instrumento de Avaliação Objetiva e Estruturada de Habilidades Técnicas Operatórias (Global Rating Scale of Objective Structured Assessment of Technical Skills)				
Cuidados com o Tecido (Respect for tissue)	1. Utilizou frequentemente de força desnecessária sobre o tecido ou causou danos ao mesmo pelo uso inapropriado dos instrumentos. (Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments.)	2	3. Manipulou cuidadosamente o tecido, mas ocasionalmente, causou danos inadvertidos. (Careful handling of tissue but occasionally caused inadvertent damage.)	4 5. Consistentemente manipulou o tecido de forma apropriada, causando danos mínimos. (Consistently handled tissues appropriately with minimal damage.)
Economia de Tempo e Movimentos (Time and motion)	1. Muitos movimentos desnecessários. (Many unnecessary moves.)	2	3. Movimentos eficientes, mas alguns desnecessários. (Efficient time/motion but some unnecessary moves)	4 5. Evidente economia de movimentos e máxima eficiência. (Economy of movement and maximum efficiency.)
Manuseio dos Instrumentos (Instrument handling)	1. Constantemente faz movimentos hesitantes ou desajeitados com os instrumentos. (Repeatedly makes tentative or awkward moves with instruments.)	2	3. Uso competente dos instrumentos, embora, ocasionalmente, apresenta-se travado ou desajeitado. (Competent use of instruments although occasionally appeared stiff or awkward.)	4 5. Movimentos ajustados e fluidos com os instrumentos. (Fluid moves with instruments and no awkwardness.)
Conhecimento dos Instrumentos (Knowledge of instruments)	1. Frequentemente usou ou solicitou instrumentos inapropriados. (Frequently asked for the wrong instrument or used an inappropriate instrument.)	2	3. Conhecia o nome da maioria dos instrumentos e os utilizou adequadamente para a tarefa. (knew the names of most instruments and use appropriate instrument for the task.)	4 5. Evidentemente familiarizado com os instrumentos requisitados e com os seus respectivos nomes. (Obviously familiar with the instruments required and their names.)
Fluxo operatório e antecipação no planejamento cirúrgico (Flow of operation and forward planning)	1. Frequentemente interrompeu o procedimento operatório ou necessitou discutir sobre o próximo passo. (Frequently stopped operating or needed to discuss next move.)	2	3. Demonstrou capacidade de antecipação no planejamento operatório com progressão contínua do procedimento. (Demonstrate ability for forward planning with steady progression of operative procedure.)	4 5. Evidentemente planejou o curso da operação, sem esforços para avançar no passo-a-passo da cirurgia. (Obviously planned course of operation with effortless flow from one move to the next.)
Uso de Auxiliares (Use of assistants)	1. Consistentemente alocou mal os auxiliares ou falhou ao utilizá-los. (Consistently placed assistants poorly or failed to use assistants.)	2	3. Bom uso dos auxiliares na maior parte do tempo. (Good use of assistants most of the time.)	4 5. Utilizou os auxiliares estrategicamente, com o máximo proveito durante todo o tempo. (Strategically used assistant to the best advantage at all times.)
Conhecimento do Procedimento Operatório Específico (Knowledge of specific procedure)	1. Conhecimento deficiente. Necessitou de instrução específica na maioria dos passos operatórios. (Deficient knowledge. Needed specific instruction at most operative steps.)	2	3. Conhecia todos os aspectos importantes da operação. (Knew all important aspects of the operation.)	4 5. Demonstrou familiaridade em todos os aspectos da operação. (Demonstrate familiarity with all aspects of the operation.)



FIGURE 1.

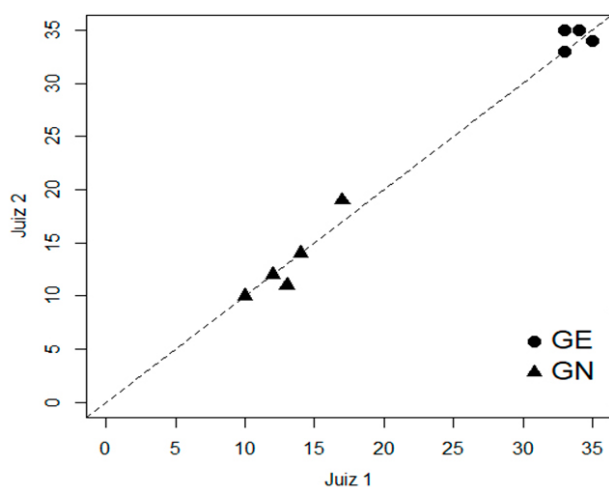


FIGURE 2.

the stage of the committee's consensual version, grammatical errors and were corrected, and some items of version T1-2 were modified by the authors. In Table 1, it is possible to see the final version of the adapted evaluation instrument.

During the pre-test, all participants found it easy to interpret and understand the instrument. Only the item "Surgery Flow and Anticipation in Surgical Planning" was not considered very clear by all surgeons. However, the three surgeons who did not give the maximum score to this item gave it a 4 (clear), which was also acceptable.

During the face validity process, in which the instrument was assessed in general, the percentage of agreement between grades 5 (very capable) among the participants was 85%. During the content validation, in which each item of the instrument was assessed individually, only the items "Knowledge of the Instruments" and "Use of Assistants" received a grade 3 (moderately capable) from one surgeon. All others receive grades 4 or 5 (capable or very capable). Still, the item "Knowledge of the Instruments" had an agreement of grades 5 of 70% and 65% for the item "Use of Assistants". The consistency between evaluators was excellent, with a Cronbach's Coefficient Alpha of 0.954.

The experts received considerably higher grades than the novices, evidencing clear discrimination between the two groups and resulting in construct validity. The mean and standard deviation of the scores assigned to experts were, respectively, 34 and 0.894 for Judge 1 and 34.33 and 0.816 for Judge 2. As for the novices, these statistics were 13.33 and 2.338 for Judge 1 and 13.33 and 3.204 for Judge 2. Figure 2 compares the scores given by each judge to each individual evaluated. The correlation coefficient (Pearson) between the two judges was 0.9944, with a CI of 95% between 0.9797 and 0.9985 and $p < 10^{-10}$, demonstrating the excellent reproducibility of the instrument.

In initial traditional subjective evaluation, the grades attributed by the judges dissented in three videos, while, in the second moment, after the use of the FV-GRS-OSATS, both judges agreed in all assessments. In addition, the first judge evaluated two videos differently in both stages. In the subjective evaluation, the judge considered that two participants should receive grade D, while in the objective assessment, their grades were C. The ratings of the second judge differed three times between the two stages, and, in the subjective evaluation, an expert was given grades equivalent to those of a novice, i.e., both received B.

DISCUSSION

This is the first study to cross-culturally adapt the Osats to Brazilian Portuguese and validate it for use in Brasil. Santos & Salles¹⁷ described the validation of specific checklists for assessing performance in some surgical procedures but did not use the global classification scale of the Osats. Few authors have published

studies in which the Osats was used in Brasil, but did not bother with its previous transcultural adaptation and validation for our country^{18,19}.

To fulfill its purposes, an instrument that was developed in one culture and will be used in another must be, first, subjected to a test of cultural equivalence^{20,21}. In addition, the assessment must have validity and reliability, which is related to the consistency and accuracy of the results of the measurement process²².

During the process of transcultural adaptation of the GRS-OSATS to Brazilian Portuguese, changes were necessary to adapt it, such as was the case with the idiomatic expression “Time and Motion”, which was adapted as “Economia de Tempo e Movimentos”. These adjustments were made because, when the committee met, it was observed that the understanding of the items varied. For this reason, the adaptation of an instrument into another language is a complex process, which cannot be done by just a simple translation. For Beaton et al.²³, the Committee plays a fundamental role in cross-cultural adaptation. In our study, the Committee evaluation showed that, although no radical change in the items was required, the adjustments provided homogenization of their understanding by the evaluators and allowed for the next step of the study: validation.

After the transcultural adaptation was completed, the face and content validity were considered appropriate for the instrument’s ability to measure what it proposes to measure. During the content validation, two items stood out for, although having achieved a high rate of agreement of grades 5, receiving smaller grades in relation to the others: “Knowledge of the Instruments” and “Use of Assistants”. In the specific case of the item “Knowledge of the Instruments”, the measurement of operative technical skills can be underestimated. Sometimes, the resident might know when and how to use the instrument, but not its name. In addition, he can also use an instrument considered inadequate and achieve the expected result.

Another item that received smaller grades was “Use of Assistants,” which might be justified by the fact that the resident assessment of the use of assistants does not depend only on their technical capabilities. It also depends on the surgical assistants themselves. In this case, for example, the measurement of the operative technical skills of residents may be overestimated if the assistants are

very good and familiar with all aspects of the operation. The proactivity of the assistants, so desired in surgical practice, may compromise the ability to assess the resident since it is not possible to know where the ability of the assessed individual begins and ends. The maintenance of the surgical team, with the same assistants in all operations, could solve this limitation.

The construct and concurrent validities were considered appropriate since the evaluations using the GRS-OSATS allowed for clear discrimination between the experts and novices and were more reliable and consistent than the subjective evaluation. Although this is the best tool researched so far, and the results of our study have proven the effectiveness of an objective evaluation in relation to a subjective one, the Judge’s personal opinion is still prevalent over the instrument, which makes its application more difficult. In this context, the judges must be trained to standardize their assessment and reduce their subjectivity.

With that regard, it is worth reflecting on what is intended with this type of instrument. Its most important role perhaps is for individual feedback since it enables residents to follow their own evolution in a systematized way. Due to the influence of context, the comparison of performance evaluations, using the GRS-OSATS, between residents is not trustworthy; therefore, this instrument should not be used for ranking. It is important that the residents’ activities are not reduced to standardized items that engender their actions. More important than measuring is improving the performance of surgeons, aiming for the development of health care.

The sampling used in the instrument validation process was of convenience in a public university hospital. The evaluation of operative performance in a single reference center and the reduced sample are the main limitations of this study.

The Brazilian Portuguese version of the GRS-OSATS maintained semantic, experimental, and conceptual equivalence with the original instrument. Face, content, construct, and concurrent validities were achieved. Thus, the instrument proved to be reproducible and reliable for use in Brasil.

ACKNOWLEDGMENTS

We thank all pregnant women who donated their placentas for the practice of surgical techniques, residents, and experts who participated in the study.

Conflicts of interest

The authors have no conflict of interest for the present study.

Contribution of the authors

Marcelo Esteves Chaves Campos (design,

structuring, planning, analysis, interpretation, writing). Marcelo Magaldi Ribeiro de Oliveira (advice, interpretation, revision). Lilian Bambirra de Assis (structuring, writing, revision). Augusto Barbosa Reis (coordination, data collection, revision). Flávio Bambirra Gonçalves (analysis, methodology, interpretation).

RESUMO

OBJETIVOS: *Objetivou-se com este trabalho adaptar transculturalmente o instrumento Objective Structured Assessment of Technical Skill (Osats) para o português-brasileiro e validá-lo no Brasil.*

MÉTODOS: *Uma versão em português-brasileiro do Osats foi criada por meio de um processo de tradução, retrotradução, versão consensual por um comitê de especialistas e pré-teste, seguido da etapa de validação. Para validades de constructo e concorrente, foram recrutados 12 participantes da Universidade Federal de Minas Gerais, divididos em um grupo de seis especialistas e um grupo de seis novatos, que tiveram de realizar tarefas em modelos de simulação utilizando placentas humanas. Cada participante foi filmado em anonimato e dois examinadores avaliaram os seus desempenhos usando o método tradicional subjetivo e depois a versão em português-brasileiro do Osats.*

RESULTADOS: *A versão em português-brasileiro do Osats alcançou as validades de face, de conteúdo, de constructo e concorrente. A média e o desvio padrão das pontuações atribuídas aos especialistas foram, respectivamente, 34 e 0,894, para o Juiz 1 e 34,33 e 0,816 para o Juiz 2. No caso dos novatos, foram 13,33 e 2,338 para o Juiz 1 e 13,33 e 3,204 para o Juiz 2. O Coeficiente de Correlação (de Pearson) entre os dois juizes foi de 0,9944 com IC 95% entre 0,9797 e 0,9985, com $p < 10^{-10}$, evidenciando a excelente reprodutibilidade do instrumento.*

CONCLUSÃO: *A versão em português-brasileiro do Osats manteve-se equivalente ao instrumento original e foi validada. Assim, pode ser usada para avaliar a performance operatória dos residentes em cirurgia no Brasil.*

PALAVRAS-CHAVE: *Avaliação educacional. Simulação. Capacitação. Treinamento. Procedimentos cirúrgicos operatórios/educação. Inquéritos e questionários.*

REFERENCES

- Larson JL, Williams RG, Ketchum J, Boehler ML, Dunnington GL. Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. *Surgery*. 2005;138(4):647-9.
- Benson A, Markwell S, Kohler TS, Tarter TH. An operative performance rating system for urology residents. *J Urol*. 2012;188(5):1877-82.
- Lipman JM, Marderstein EL, Zeinali F, Phitayakorn R, Ponsky JL, Delaney CP. Objective evaluation of the performance of surgical trainees on a porcine model of open colectomy. *Br J Surg*. 2010;97(3):391-5.
- Herbella FAM, Velanovich V. Observations on multi-generational interactions in academic surgical practice and education. *Rev Assoc Med Bras* (1992). 2019;65(2):105-9.
- Procianoy RS, Silveira RC. Objective structured clinical assessment as an evaluation tool for medical students. *Rev Assoc Med Bras* (1992). 2009;55(6):632-3.
- Lafracia FM, Herbella FAM, Kalluf JR, Schlottmann F, Patti MG. Attitudes and experiences during training and professional expectations in generation-y surgical residents. *Rev Assoc Med Bras* (1992). 2019;65(3):348-54.
- Shaharan S, Neary P. Evaluation of surgical training in the era of simulation. *World J Gastrointest Endosc*. 2014;6(9):436-47.
- Atesok K, Satava RM, Marsh JL, Hurwitz SR. Measuring surgical skills in simulation-based training. *J Am Acad Orthop Surg*. 2017;25(10):665-72.
- Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. *Acad Med*. 1996;71(12):1363-5.
- Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273-8.
- Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. *Am J Surg*. 1997;173(3):226-30.
- Kara CO, Mengi E, Tümkaya F, Ardiç FN, Şenol H. Adaptation of "Objective Structured Assessment of Technical Skills" for adenotonsillectomy into Turkish: a validity and reliability study. *Turk Arch Otorinolaryngol*. 2019;57(1):7-13.
- Duarte PS, Miyazaki MC, Ciconelli RM, Sesso R. Translation and cultural adaptation of the quality of life assessment instrument for chronic renal patients (KDQOL-SF). *Rev Assoc Med Bras* (1992). 2003;49(4):375-81.
- Stevellink SA, van Brakel WH. The cross-cultural equivalence of participation instruments: a systematic review. *Disabil Rehabil*. 2013;35(15):1256-68.
- Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46(12):1417-32.
- Oliveira Magaldi M, Nicolato A, Godinho JV, Santos M, Prosdociami A, Malheiros JA, et al. Human placenta aneurysm model for training neurosurgeons in vascular microsurgery. *Neurosurgery*. 2014;10(Suppl 4):592-600.
- Santos EG, Salles GF. Construction and validation of a surgical skills assessment tool for general surgery residency program. *Rev Col Bras Cir*. 2015;42(6):407-12.
- Barreira MA, Siveira DG, Rocha HA, Moura Junior LG, Mesquita CJ, Borges GC. Model for simulated training of laparoscopic gastroenterostomy. *Acta Cir Bras*. 2017;32(1):81-9.
- Tube MI, Spencer-Netto FA, Oliveira AI, Holanda AC, Barros BL, Rezende CC, et al. Surgical model pig ex vivo for venous dissection teaching in medical schools. *Acta Cir Bras*. 2017;32(2):157-67.
- Silveira C, Parpinelli MA, Pacagnella RC, Camargo RS, Costa ML, Zanardi DM, et al. Cross-cultural adaptation of the World Health Organization Disability Assessment Schedule (WHODAS 2.0) into Portuguese. *Rev Assoc Med Bras* (1992). 2013;59(3):234-40.
- Soárez PC, Ciconelli RM, Pavin T, Ogata AJ, Curci KA, Oliveira MR. Cross-cultural adaptation of the CDC Worksite Health ScoreCard questionnaire into Portuguese. *Rev Assoc Med Bras* (1992). 2016;62(3):236-42.
- Araújo RB, Fortes MR, Abbade LP, Miot HA. Translation, cultural adaptation to Brasil and validation of the Venous Leg Ulcer Quality of Life Questionnaire (VLU-QoL-Br). *Rev Assoc Med Bras* (1992). 2014;60(3):249-54.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* (Phila Pa 1976). 2000;25(24):3186-91.

