

# Assessment of cross-cultural adaptations and measurement properties of self-report outcome measures in Portuguese relevant to temporomandibular disorders: a systematic review

*Avaliação das adaptações transculturais e propriedades de medida de questionários em língua portuguesa relacionados às desordens temporomandibulares: uma revisão sistemática*

*Evaluación de las adaptaciones transculturales y propiedades de medida de cuestionarios en lengua portuguesa relacionados con trastornos temporomandibulares: una revisión sistemática*

Leticia Miranda Resende da Costa<sup>1</sup>, Daiane Lazzeri de Medeiros<sup>2</sup>, Lilian Gerdi Kittel Ries<sup>2</sup>, Audria Beretta<sup>2</sup>, Marcos Amaral de Noronha<sup>2,3</sup>

**ABSTRACT** | The aim of this study was to identify, through a systematic review, which questionnaires used to assess temporomandibular (TMD) disorders are available in the Portuguese language, describing and analyzing the procedures used to translate and adapt the questionnaire into Portuguese, as well their measurement properties. Systematic searches were performed in five electronic databases (MEDLINE by PubMed, Embase, CINAHL by EBSCO, SciELO and LILACS). All studies were analyzed according to the criteria of quality guidelines for the procedures for cross-cultural adaptation and measurement properties. There were 1418, of which only six were eligible. Manual search retrieved two additional articles and a book chapter, totaling nine eligible studies. Four instruments were translated/adapted into Portuguese: The Research Diagnostic Criteria for Temporomandibular Disorders: Axis II (RDC/TMD); the Fonseca Questionnaire and Anamnestic Index; the Questionnaire from the American Academy of Orofacial Pain (QAADO); and Mandibular Function Impairment Questionnaire (MFIQ), but none of these questionnaires were fully submitted to the steps for translation and adaptation recommended by the followed guidelines and none

of the questionnaires had all measurement properties tested. The RDC/TMD and the MFIQ are the best available questionnaires in Portuguese to assess TMD, as were those with the most measurement properties acceptable results tested.

**Keywords** | Temporomandibular Joint; Questionnaires; Stomatognathic System.

**RESUMO** | O objetivo deste estudo foi identificar, por meio de uma revisão sistemática, os questionários que avaliam as desordens temporomandibulares (DTM) disponíveis em português, bem como descrever e analisar os procedimentos de tradução e adaptação para a língua portuguesa dos mesmos e suas respectivas propriedades de medida. Foram realizadas buscas sistematizadas em cinco bases de dados (MEDLINE via PubMed, Embase, CINAHL via EBSCO, SciELO e LILACS). Todos os estudos foram analisados de acordo com os critérios de qualidade de diretrizes para os procedimentos de adaptação transcultural e propriedades de medida. Foram encontrados 1418 estudos, sendo que apenas 6 foram considerados elegíveis. Foram incluídos em busca manual dois artigos e um capítulo de livro, totalizando nove estudos. Quatro instrumentos foram traduzidos/

Study conducted at the *Laboratório de Desenvolvimento e Controle Postural (LADESCOP)* of the *Centro de Ciências da Saúde e do Esporte* of the *Universidade do Estado de Santa Catarina (CEFID-UDESC)* - Florianópolis (SC), Brazil.

<sup>1</sup>Graduate Program in Sciences of the Human Movement of the School of Physical Education (ESEF) of the Universidade Federal do Rio Grande do Sul (UFRGS) - Porto Alegre (RS), Brazil.

<sup>2</sup>Graduate Program in Physical Therapy of the CEFID-UDESC - Florianópolis (SC), Brazil.

<sup>3</sup>La Trobe Rural Health School, Allied Health Department - Bendigo, Victoria, Australia.

adaptados para a língua portuguesa: *Research Diagnostic Criteria for Temporomandibular Disorders* (RDC/TMD), Questionário e Índice Anamnésico de Fonseca, Questionário da Academia Americana de Dor Orofacial (QAADO) e *Mandibular Function Impairment Questionnaire* (MFIQ), porém nenhum desses realizou completamente as etapas de tradução ou análise das propriedades de medida. O RDC/TMD e o MFIQ foram considerados os instrumentos mais adequados, pois foram os que apresentaram mais propriedades de medida apropriadamente testadas.

**Descritores** | Articulação Temporomandibular; Questionários; Sistema Estomatognático.

**RESUMEN** | El objetivo de este estudio fue identificar, por medio de una revisión sistemática, los cuestionarios que evalúan los trastornos temporomandibulares (TTM) disponible en portugués, además de describir y analizar los procedimientos de traducción y adaptación al portugués de los mismos y sus respectivas propiedades de medida. Fueron realizadas búsquedas sistemáticas en cinco bases de

datos (MEDLINE vía PubMed, Embase, CINAHL a través de EBSCO, SciELO y LILACS). Se analizaron todos los estudios de acuerdo a los criterios de calidad de directrices para los procedimientos de adaptación transcultural y propiedades de medida. Fueron encontrados 1.418 estudios, siendo que solo 6 fueron considerados elegibles. Fueron incluidos en búsqueda manual dos artículos y un capítulo de libro, totalizando nueve estudios. Cuatro instrumentos fueron traducidos/adaptados al portugués: *Research Diagnostic Criteria for Temporomandibular Disorders* (RDC/TMD), Questionário e Índice Anamnésico de Fonseca, Questionário da Academia Americana de Dor Orofacial (QAADO) y *Mandibular Function Impairment Questionnaire* (MFIQ), pero ninguno de estos llevó a cabo plenamente las etapas de traducción y análisis de las propiedades de medida. El RDC/TMD y el MFIQ fueron considerados los instrumentos más adecuados, ya que fueron los que presentaron más propiedades de medida apropriadamente examinadas.

**Palabras clave** | Articulación tempomandibular; Cuestionarios; Sistema Estomatognático.

## INTRODUCTION

Temporomandibular disorder (TMD) is the term applied to the functional alterations related to the temporomandibular joint (TMJ) and the associated masticatory structures<sup>1</sup>. These changes are mainly characterized by pain in the temporomandibular region or in the muscles of mastication, limitations or deviations in the movement of the mandible, and sounds from the TMJ during jaw functioning<sup>2</sup>.

The etiology of the TMD is multifactorial<sup>3</sup>, with special attention to: malocclusion, mandibular trauma or in the TMJ, muscle alterations, parafunctional habits<sup>1</sup>, emotional or postural problems<sup>4</sup>.

Several instruments for the diagnosis of TMD have been presented, however there is no consensus<sup>5</sup>. Advantages, disadvantages and limitations, as well as distinct applicability, are observed<sup>6</sup>. One limitation is the fact that most of them have been developed in English. In order for these instruments to be used in Brazil, the cultural adaptation and the testing of their measurement properties are recommended<sup>7-9</sup>.

There are questionnaires related to TMD which have been adapted into Portuguese and had some of their measurement properties evaluated and tested, but there is no study which summarizes this information though. Thus, the aim of this study was to describe and analyze, through a systematic review, the procedures of their translation and adaptation into Portuguese, as

well as their measurement properties of questionnaires which assessed the TMDs, verifying then which the best available questionnaires for TMD in Brazil are.

## METHODOLOGY

In order to identify the instruments which assess TMD in Portuguese, systematic searches in five databases (MEDLINE by PubMed, Embase, CINAHL by EBSCO, SciELO and LILACS) were performed. The search terms and the operators used were: (stomatognathic OR temporomandibular disorders OR tmd OR oro-facial pain OR orofacial pain OR craniomandibular disorders OR chewing OR masticatory muscles OR temporomandibular joint OR tmj) AND (questionnaire OR index OR scale OR score OR assessment OR evaluation OR self-report OR inventory) AND (Brazil OR Brasil OR Portuguese OR Brazilian Portuguese OR Brazilian). The searches were not limited to language or publication date. The last search was conducted in January 2013.

Studies which reported questionnaires developed to assess the TMD and cannot be combined with dysfunction of other joints or muscles which are not part of this system, and which have been subjected to any procedure of translation and/or adaptation and validation into Portuguese were included, in which their measurement properties were tested.

The selection of articles was performed by three independent evaluators; disagreements were solved by consensus. The selection was initially performed by titles, followed by abstracts and then, finally, by the full reading of the article.

The included studies were assessed according to the guidelines for the procedures for cross-cultural adaptation of questionnaires<sup>7</sup>, adapted by Menezes-Costa<sup>10</sup>, which comprises the following steps: translation, synthesis of the translations, back translation, committee review and pretesting of the final version. The classification of each item was scored as positive (+): procedure performed according to the guidelines of the quality criteria used; negative (-): procedure performed as expected, though with insufficient number of translators and/or back-translators; null (0): no information was available in order to quantify the stage; or undefined (?): design or method not appropriated or unspecified. Further details about the stages of translation can be found in the article of Beaton *et al.*<sup>7</sup>.

According to the guidelines used, the translation should be performed by two or more translators independently. In the synthesis of the translations, translators synthesize all translations and produce a consensus version. The back translation must be performed by translators who have no knowledge of the questionnaire and who translate the consensual version to the original language of the questionnaire. The analysis of the committee should be performed by specialists who examine all versions of the questionnaire and develop the pre-final version. This version is tested on members of the target population and it is ended with a pretest.

For the analysis of the measurement properties of the selected articles, the quality criteria for measurement properties of questionnaires in the Health Area<sup>8</sup>, adapted by Menezes-Costa<sup>10</sup>, were used. The following aspects were analyzed: internal consistency, construct validity, reproducibility (concordance and reliability), responsiveness and Ceiling and Floor effects.

The quality of each step of the quality criteria for measurement properties were classified as positive (+): procedures performed in accordance with the quality criteria adopted; undefined (?): questionable methods or design of the study; negative (-): data referring to each property had values greater or lower than those defined by the criteria adopted in the study; or null (0): there was not enough information to qualify each measurement property.

According to the criteria of quality, internal consistency should be measured by Cronbach's alpha, which indicates the level of homogeneity of the instrument or of its consisting questions. This statistical tool is used to

verify whether the exclusion of a particular issue affects the consistency of the instrument. The recommended value for Cronbach's alpha is between 0.70 and 0.95<sup>8</sup>.

The construct validity examines the extent of each score of a specific questionnaire related to other similar measures, so that the hypotheses are strongly related to the concepts being measured, of which at least 75% of the results shall be in accordance with such hypotheses.

Reproducibility is the degree in which repeated measures in the same individual results in similar answers. Reproducibility includes reliability and concordance. As to reliability, it is considered positive when the Intraclass Correlation Coefficient (ICC) or kappa is  $\geq 0.70$ .

The concordance measures how much two or more measures are close to each other. The positive rating occurs when Minimally Important Change (MIC) > Minimum Detectable Change (MDC) or MIC outside the Concordance Limits (CL), or other convincing arguments.

Response is the ability of the survey to detect clinical changes over time. It was considered positive when the individual MDC or the group MDC < MIC or the MIC was outside the CL, or the responsivity reason > 0.96, or the area under the curve  $\geq 0.70$ .

The effects of Ceiling and Floor are defined by the number of respondents who achieved the maximum or the minimum possible score; in order to obtain a positive score, less than 15% of the respondents should reach either the maximum or the minimum possible score.

## RESULTS

There were found 1,418 studies, of which only 6 were eligible for data analysis. Two articles and one book chapter (Figure 1) were included by manual search, totaling nine eligible studies. Four different instruments were translated/adapted into Portuguese, namely: Research Diagnostic Criteria for Temporomandibular Disorders: Axis II (RDC/TMD)<sup>11-14</sup>, Questionnaire and Anamnestic Index of Fonseca<sup>5,15</sup>, Questionnaire of the American Academy of Orofacial Pain (QAAOP)<sup>2,16</sup> and Mandibular Function Impairment Questionnaire (MFIQ)<sup>17</sup> (Table 1).

Among the nine studies analyzed, four translated and/or adapted an instrument. None fully performed all stages of cultural adaptation according to the guidelines of Beaton *et al.*<sup>7</sup>. The steps of synthesis, committee review and pretesting were properly performed for

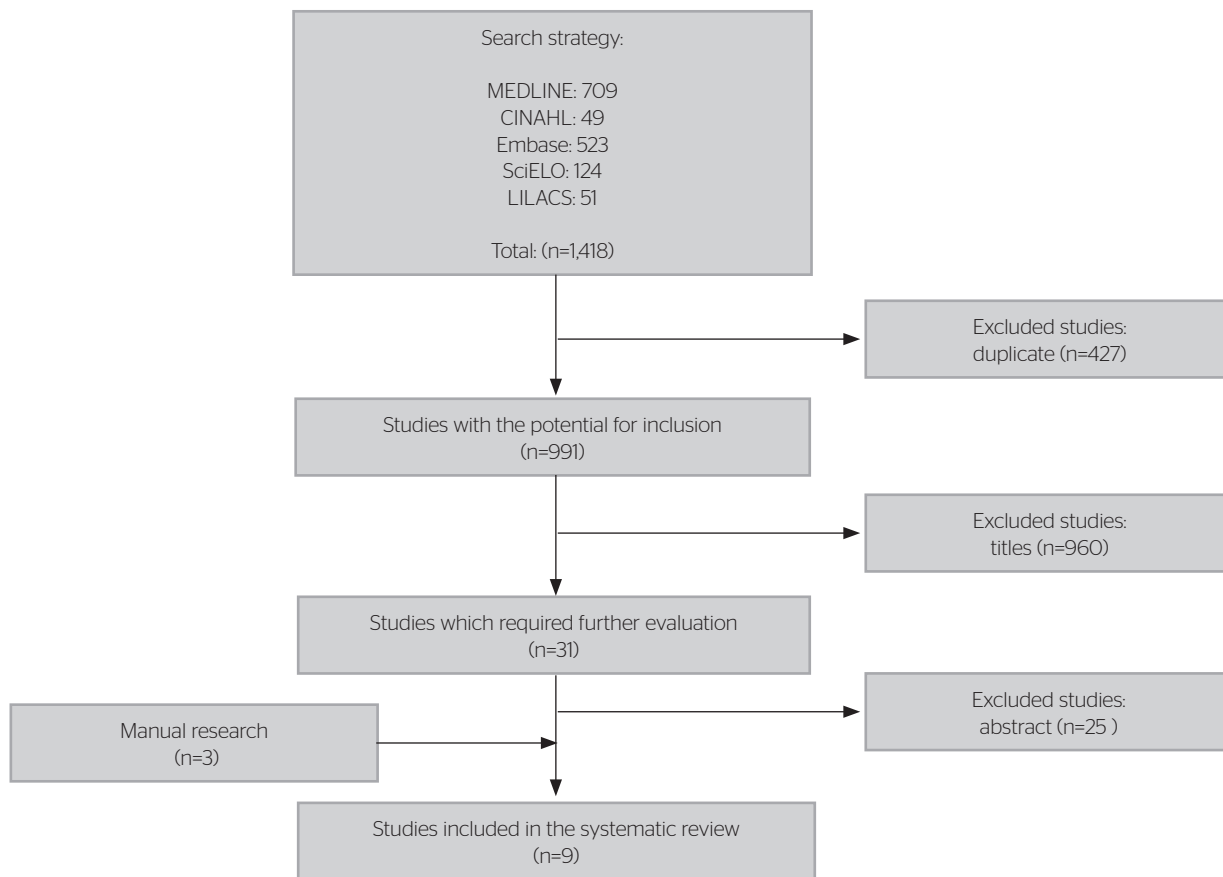


Figure 1. Flowchart of the systematic review

Table 1. Analysis of the questionnaires in Portuguese for evaluation of temporomandibular disorders according to the guidelines of the process of cross-cultural adaptation of questionnaires<sup>9</sup>

Study	Translation	Synthesis	Back Translation	Committee Review	Pre-test
QAAOP <sup>2</sup>	0	0	0	0	0
RDC/TMD <sup>11</sup>	0	+	0	+	+
FONSECA <sup>15</sup>	0	0	0	0	0
MFIQ <sup>17</sup>	+	+	-	+	+

RDC/TMD: Research Diagnostic Criteria for Temporomandibular Disorders: Axis II; FONSECA: Questionnaire and Anamnestic Index of Fonseca; QAAOP: Questionnaire of the American Academy of Orofacial Pain; MFIQ: Mandibular Function Impairment Questionnaire; +: positive classification; -: negative classification; 0: no information

the RDC instrument. The QAAOP was translated by a single translator, the Questionnaire and Anamnestic Index of Fonseca was adapted from an existing questionnaire<sup>18</sup>, however, there is no information on how the cultural adaptation was conducted, and the MFIQ<sup>17</sup> was adequately translated, the synthesis, the analysis and the pre-test were performed, but the back-translation was carried out by a translator.

The results of the evaluations of selected studies according to the quality criteria for measurement properties of questionnaires in the Health Area<sup>8</sup> are shown in Table 2.

Regarding the measurement properties, in the study of Fonseca<sup>15</sup> the reliability was analyzed through the Spearman correlation test and the simple regression of linear analysis, in which the reliability of the proposed values for the anamnesis and clinical examination was approximately 95%. The construct validity in the study of Lucena<sup>14</sup> references to this analysis, however, as the hypotheses were not specified, this question was not considered valid.

## DISCUSSION

The adaptation of an instrument is required when there is a distinction between the culture of the place of preparation and where the resource will be used, as there may be a disagreement on the interpretation of the questionnaire's items<sup>19</sup>. Thus, it is necessary that an accurate assessment of its cultural adjustment is made, as well as its measurement properties<sup>20</sup>.

Most steps of the cross-cultural adaptation of the RDC/TMD<sup>11</sup> and the MFIQ<sup>17</sup> instruments were performed according to the standards adopted in this study<sup>7</sup>.

Table 2. Assessment of the measurement properties of the questionnaires in Portuguese for evaluation of temporomandibular disorders according to the quality criteria for measurement properties of questionnaires in the healthcare area<sup>10</sup>

Studies	Reproducibility (Concordance)	Reproducibility (Reliability)	Internal Consistency	Responsiveness	Construct's validity	Ceiling and Floor effects
FONSECA <sup>5</sup>	0	?	0	0	0	0
QAAOP <sup>16</sup>	-	0	0	0	0	0
RDC/TMD <sup>14</sup>	0	+	-	0	?	0
RDC/TMD <sup>13</sup>	+	0	+	0	0	0
MFIQ <sup>17</sup>	0	+	+	0	+	0
FONSECA <sup>5</sup>	-	+	?	0	0	0
RDC/TMD <sup>12</sup>	+	0	+	0	0	0

RDC/TMD: Research Diagnostic Criteria for Temporomandibular Disorders: Axis II; FONSECA: Questionnaire and Anamnestic Index of Fonseca; QAAOP: Questionnaire of the American Academy of Orofacial Pain; MFIQ: Mandibular Function Impairment Questionnaire; +: positive classification; -: negative classification; 0: unavailable information; ?: undefined design or method

The RDC/TMD<sup>11</sup> had the steps of synthesis, committee review and pretesting. In the MFIQ<sup>17</sup>, the translation, the synthesis, the analysis and the pre-test were performed; however, the back-translation was done by a single translator. The QAAOP was translated into Portuguese and presented in a book chapter<sup>2</sup>, without a description of the process. Now, the Questionnaire and Anamnestic Index of Fonseca<sup>15</sup> was built in Brazil, in Portuguese, and mostly based on the studies of Helkimo<sup>21,22</sup>.

None of the questionnaires completely evaluated all measurement properties. The RDC/TMD presented the highest number of properties tested so far (concordance, reliability, internal consistency and construct validity), followed by the MFIQ, which evaluated reliability, internal consistency and construct validity.

In this review, it was possible to identify that most that most of the questionnaires had their concordance assessed, and only two of them<sup>12,13</sup> showed adequate results, however, both of them assessed the RDC/TMD (Table 2). Two different studies have also examined the concordance: one assessed the QAAOP<sup>16</sup>; the other, the Questionnaire and Anamnestic Index of Fonseca<sup>5</sup>, but had minimally important change within the limits of concordance.

The reliability was tested on three instruments, the Questionnaire and Anamnestic Index of Fonseca<sup>5</sup>, on the RDC/TMD<sup>14</sup> and the MFIQ<sup>17</sup>, since tests were applied appropriately.

The internal consistency was assessed on the RDC/TMD<sup>12,13</sup> and the MFIQ<sup>17</sup>, through the Cronbach's alpha, with results considered positive. Another study which evaluated the internal consistency of the RDC/TMD<sup>14</sup> has obtained a Cronbach's alpha ranging from 0.68 to 0.73, which was not considered to be appropriate. The study which evaluated the Questionnaire and Anamnestic Index of Fonseca<sup>5</sup> calculated the internal consistency through the Kuder-Richardson coefficient,

which was considered to be undefined. The responsiveness and the floor and ceiling effect were not tested in either study.

The construct validity was tested in two instruments, the RDC/TMD<sup>14</sup> and the MFIQ<sup>17</sup>. The study which evaluated the RDC/TMD<sup>14</sup> used the Spearman correlation test, relating the questionnaire with other similar measures. The test showed a positive correlation for the measures compared, however, no hypotheses were formulated before testing the validity of the construct, which made the design of the study undefined. In the study which evaluated the MFIQ<sup>17</sup>, the construct's validity was considered to be positive.

It is important to note that the Questionnaire and Index Anamnesis of Fonseca is considered to be simple and easy to apply, and, therefore, their use has been favored in epidemiological population studies, though it does not offer a TMD diagnostic classification<sup>3</sup>.

The MFIQ has a scoring system which allows classifying categories of severity of functional limitation related to TMD<sup>3</sup>. One of its advantages is to measure the functional limitation on the TDM, distinguishing it from other indexes which assess the severity of the signs and symptoms<sup>3</sup>. However, such questionnaire is not well distributed worldwide, which can be observed in this study, where we observed only one manuscript addressing the theme.

Using the QAAOP associated to a review of the history and clinical examination is recommended<sup>2</sup>. Furthermore, in the reported study, the authors found that the QAAOP is useful for pre-screening patients, but it does not allow the obtaining of the diagnosis<sup>16</sup>.

The RDC/TMD has been translated into 18 languages, including Portuguese<sup>23</sup>, and for this reason it is widely used, however, it is fairly complete and extensive. Because it is a criterion for research, some patients with signs and symptoms of TMD may not be considered as bearers of such according to the RDC/TMD ratings<sup>6</sup>. Nevertheless,

this is an instrument which includes a large proportion of the aspects involving the TMDs, transforming it in a good and well accepted diagnostic tool in literature.

Although the searches occur in the most used databases, some studies have not been selected and it was opted for a manual search. Still, some studies may not have been included, since some Brazilian journals might not be indexed in the databases used, this being considered as a limiting factor in this review.

## CONCLUSION

The RDC/TMD and the MFIQ may be considered the most suitable instruments to be used when compared to other tested questionnaires, as they were the ones which had more appropriately tested stages of cultural adaptation and properties so far.

It is noteworthy that this study may guide the performing of new evaluative studies on measurement properties which have not yet been tested or implemented according to the guidelines used.

Since the DRC is a widely used and widespread tool in many countries, it is suggested that an adaptation of the RDC/TMD, focused on clinical practice, is considered, as this instrument was developed for research purposes, being considered long and difficult to be applied; however, it is a good instrument for classifying the TMDs and its severity.

## REFERENCES

- American Society of Temporomandibular Joint Surgeons. Guidelines for diagnosis and management of disorders involving the temporomandibular joint and related musculoskeletal structures. *Cranio*. 2003;21(1):68-76.
- Okeson JP. *Dor Orofacial - Guia para Avaliação, Diagnóstico e Tratamento*. Academia Americana de Dor Orofacial. São Paulo: Quintessence; 1998.
- Chaves TC, Oliveira AS, Grossi DB. Principais instrumentos para avaliação da disfunção temporomandibular, parte I: índices e questionários; uma contribuição para a prática clínica e de pesquisa. *Fisioter Pesqui*. 2008;15(1):92-100.
- Oliveira AS, Dias EM, Contato RG, Berzin F. Prevalence study of signs and symptoms of temporomandibular disorder in Brazilian college students. *Braz Oral Res*. 2006;20(1):3-7.
- Campos JADB, Gonçalves DAG, Camparis CM, Speciali JG. Confiabilidade de um formulário para diagnóstico da severidade da disfunção temporomandibular. *Rev Bras Fisioter*. 2009;13(1):38-43.
- Chaves TC, Oliveira AS, Grossi DB. Principais instrumentos para avaliação da disfunção temporomandibular, parte II: critérios diagnósticos; uma contribuição para a prática clínica e de pesquisa. *Fisioter Pesqui*. 2008;15(1):132.
- 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.
- Terwee CB, Bot SD, Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42.
- Maher CG, Latimer J, Costa LOP. The relevance of cross-cultural adaptation and clinimetrics for physical therapy instruments. *Rev Bras Fisioter*. 2007;11(4):245-52.
- Menezes Costa LC, Maher CG, McAuley JH, Costa LO. Systematic review of cross-cultural adaptations of McGill Pain Questionnaire reveals a paucity of clinimetric testing. *J Clin Epidemiol*. 2009;62(9):934-43.
- Kosminsky M, Lucena LBS, Siqueira JTT, Pereira Júnior FJ, Góes, PSA. Adaptação cultural do questionário "Research diagnostic criteria for temporomandibular disorders: axis II" para o português. *JBC J Bras Clin Odontol Integr*. 2004;8(43):51-61.
- Campos JADB, Carrascosa AC, Loffredo LCM, Faria JB. Consistência interna e reprodutibilidade da versão em português do critério de diagnóstico na pesquisa para distúrbios temporomandibulares (RDC/TMD - Eixo II). *Rev Bras Fisioter*. 2007;11(6):451-9.
- Cavalcanti RF, Studart LM, Kosminsky M, Goes PSA. Validation of the multimedia version of the RDC/TMD axis II questionnaire in Portuguese. *J Appl Oral Sci*. 2010;18(3):231-6.
- Lucena LBS, Kosminsky M, Costa LJ, Goes PSA. Validation of the Portuguese version of the RDC/TMD Axis II questionnaire. *Braz Oral Res*. 2006;20(4):312-17.
- Fonseca DM, Bonfante G, Valle AL, Freitas SFT. Diagnóstico pela anamnese da disfunção craniomandibular. *Rev Gaúcha Odontol*. 1994;42(1):23-8.
- Manfredi APS, Silva AAD, Vendite LL. The sensibility appreciation of the questionnaire for selection of orofacial pain and temporomandibular disorders recommended by the American Academy of Orofacial Pain. *Rev Bras Otorrinolaringol*. 2001;67(6):763-8.
- Campos JA, Carrascosa AC, Maroco J. Validity and reliability of the Portuguese version of Mandibular Function Impairment Questionnaire. *J Oral Rehabil*. 2012;39(5):377-83.
- Helkimo M. Studies on function and dysfunction of the masticatory system. IV. Age and sex distribution of symptoms of dysfunction of the masticatory system in Lapps in the north of Finland. *Acta Odontol Scand* 1974;32(4):255-67.
- Moser AMM, Traebert J. Transcultural adaptation of the HIV/AIDS-Target Quality of Life for HIV/AIDS for the evaluation of patients' quality of life. *Ciênc Saúde Coletiva*. 2011;16(Suppl 1):1357-62.
- Lage LV, Levy RA, Ciconelli RM. Instrumentos de avaliação em reumatologia: importância de sua tradução e validação para nosso idioma. *Rev Bras Reumatol*. 2006;46(4):237.
- Helkimo M. Studies on function and dysfunction of the masticatory system. II. Index for anamnestic and clinical dysfunction and occlusal state. *Sven Tandlak Tidskr*. 1974;67(2):101-21.
- Helkimo M. Studies on function and dysfunction of the masticatory system. 3. Analyses of anamnestic and clinical recordings of dysfunction with the aid of indices. *Sven Tandlak Tidskr*. 1974;67(3):165-81.
- John MT, Dworkin SF, Mancl LA. Reliability of clinical temporomandibular disorder diagnoses. *Pain*. 2005;118(1-2):61-9.