Journal of Oral Rehabilitation

Journal of Oral Rehabilitation 2012 **39**; 377–383

Validity and reliability of the Portuguese version of Mandibular Function Impairment Questionnaire

J. A. D. B. CAMPOS*, A. C. CARRASCOSA† & J. MAROCO‡ *Universidade Estadual Paulista (UNESP), Faculdade de Odontologia de Araraquara-SP, †Universidade Estadual Paulista (UNESP), Faculdade de Ciências Farmacêuticas de Araraquara-SP, Brazil and ‡ISPA, Instituto Universitário, Departamento de Estatística e Unidade de Investigação em Psicologia e Saúde, Lisboa, Portugal

SUMMARY **The aim of this study is to evaluate the validity and reliability of the Mandibular Function Impairment Questionnaire (MFIQ) (Portuguese version). Face and content validity of the Portuguese version were performed. To assess reproducibility of the data gathered with MFIQ, it was applied to 62 individuals who completed the questionnaire on two occasions. Validity and reliability of the data gathered with MFIQ were evaluated in a sample of 249 patients. Construct-related validity was assessed through factorial validity (by means of a confirmatory factor analysis), and convergent and discriminant validities were assessed, respectively, by the average variance extracted (AVE), composite reliability (CC) and bivariate correlations between factors. The internal consistency was estimated by the standardised Cronbach's alpha coefficient (α) and reproducibility by the intra-class correlation coefficient (ICC). All the items of MFIQ showed content validity. Reproducibility was excellent in both the 'functional capacity' dimension (D1) (ICC_{D1} = 0·895, 95% CI = 0·832 to 0·935) and the 'feeding' dimension (D2) (ICC_{D2} = 0·825, 95% CI = 0·726 to 0·891). Items 1, 2, 6 and 7 of D1 had factor weights below the desired cut-off (0·5), and overall fit of the original bifactorial structure of the MFIQ was poor [(confirmatory fit index) CFI = 0·850, (goodness of fit index) GFI = 0·781, (root mean square error of approximation) RMSEA = 0·118]. Thus, these items were excluded, and the new, reduced version of the MFIQ showed good fit (CFI = 0·933, GFI = 0·879, RMSEA = 0·099). The convergent validity was adequate (AVE ≥ 0·5, CC ≥ 0·7) for both factors. However, their discriminant validity was low (AVE_{D1} = 0·51 and AVE_{D2} = 0·66 < ρ^2_{D1D2} = 0·70). The internal consistency was excellent (α_{D1} = 0·874; α_{D2} = 0·918). The Portuguese version of the reduced MFIQ produced data with good validity and reliability.**

KEYWORDS: reliability, validity, Mandibular Function Impairment Questionnaire, temporomandibular joint disorders

Accepted for publication 29 October 2011

Introduction

A variety of clinical problems related to the masticatory muscles, temporomandibular joints and associated structures known collectively as temporomandibular disorders (TMD) are, according to Okeson (1), the most frequent conditions that trigger pain of musculoskeletal origin in the masticatory system.

The pain is often associated with other signs and symptoms. These include limitations or deviations in jaw movement and sounds during mandibular function (2), feeding and psychological and social functioning, which are likely to impact negatively the quality of life of individuals with TMD (3–5). The psychological aspects (6) and limitations of mandibular function (7) have been frequently cited as comorbid with TMD.

Mandibular limitations evaluated in patients with TMD are related to the mechanical functions (opening of the mouth, chewing ability), which have social implications (talk, smile) and consequences (appearance,

doi: 10.1111/j.1365-2842.2011.02276.x

communication). Owing to the need to evaluate the full impact of TMD on the quality of life, several evaluation methods have been developed. These are especially useful in epidemiological and clinical studies, because they allow the screening of individuals in need of treatment.

Among the instruments most used in the Portuguese language to assess TMD are the questionnaire proposed by the American Academy of Orofacial Pain (1), the Index of Anamnesic (8), the Mandibular Function Impairment Questionnaire (MFIQ) (9) and the Research Diagnostic Criteria for Temporomandibular Disorders (10). The MFIQ (9) allows the classification of individuals in relation to the severity of the functional limitation, related to TMD. The MFIQ has been under use (11, 12), and the strong association of its scores with measures of pain, restricted jaw movements and psychological changes have pointed to a reliable and valid additional tool to assess the limitations of mandibular function in patients with TMD (9). However, the MFIQ was originally proposed in the English language, and as far as we know, no transcultural adaptations to other languages, with evaluation of the metric properties of the data gathered, have been proposed. Chaves *et al.* (13) made a first Portuguese unofficial translation; however, its metric qualities were not evaluated. According to Guillemin *et al.* (14) and Beaton *et al.* (15), cultural adaptation is a process that involves the combination of a component of literal translation of words and phrases from one language to another, and a meticulous process of attunement that addresses the cultural context and lifestyle of the target population. As noted by Beaton *et al.* (15), the use of instruments without their transcultural upgrade may jeopardise the validity and reliability of data gathered and conclusions reached from the analysis of that data.

With this concern, we carried out the transcultural adaptation of the 'MFIQ' for the Portuguese language and studied the reliability and validity of the data gathered with this instrument.

Materials and methods

Participants

The participants were 249 subjects who attended, from February 2009 to March 2010, the Physiotherapy Clinic of the University Center of Araraquara – UNIARA and agreed to participate, voluntary, in this study. These patients were complaining from pain or discomfort in the temporomandibular joint and were diagnosed based on RDC/TMD Axis I classification criteria. The mean age of the participants was 36·84 ± 8·95 years and 53·73% were women.

Instrument

We used the original version of the MFIQ, developed in English by Stegenga *et al.* (9). This measuring instrument consists of 17 Likert-type questions, anchored in five points ranging from '0-no difficulty' to 'four-very difficult or impossible without help'. These 17 questions are arranged in two dimensions (D1: Functional Capacity, D2: Feeding). The average of points assigned to each question allows the classification of individuals according to the TMD severity.

To perform the translation into Portuguese, the English original instrument was translated by three bilingual Brazilian translators working in the field of Dentistry and temporomandibular dysfunction. The three versions were compared, and from them, a final version was drawn up by the research team. The final Portuguese version was then given to native English translator who performed the back translation, from which the equivalence of forms was evaluated.

Face validity

The face validation process involved six dentistry professionals (specialists on temporomandibular disorders) and three experts of the English language. The idiomatic, semantic, cultural and conceptual equivalence of the instrument was analysed to obtain agreement and consensus. Thereafter, a preliminary version of the instrument was pretested in a group of 25 subjects, undergoing treatment in the Physiotherapy Clinic, in the University Center of Araraquara – UNIARA. A comprehension index (CI) was obtained for each item. Further analysis proceeded only after CI reached 80%.

Content validity

Content-related validity was assessed by 21 dentists with expertise in temporomandibular disorders. These 'judges' evaluated each of the MFIQs items and classified them as 'essential', 'useful, but non-essential' or 'not necessary'. The content validity ratio (CVR) was calculated and classified according to Laewshe (16).

Construct validity

Construct-related validity is supported by simultaneous demonstration of factorial validity, convergent validity and discriminant validity as described below.

Factorial validity

The two-factor originally proposed structured was evaluated in the sample of 249 subjects described in the *participants* section. Initially, data obtained in this stage were evaluated for psychometric sensitivity using shape (skewness and kurtosis) and central tendency descriptive statistics. Psychometric sensitivity was accepted for skewness and kurtosis absolute values smaller than three and seven, respectively [see (17)]. We conducted confirmatory factor analysis to determine the degree to which the dimensions found satisfy the expected structure. The indices χ^2/df (ratio chi-square and degrees of freedom), CFI, goodness of fit index (GFI) and root mean square error of approximation (RMSEA) with reference values for good fit given by Maroco and Byrne among others (17–19) were used to evaluate the factors' goodness of fit. Analyses were performed using the program*.

To compare the two-factor model proposed by Stegenga *et al.* (9), with the one-factor model, we first calculated the difference between the models' chi-squares and then the Akaike Information Criterion (AIC), Browne-Cudeck Criterion (BCC) and Bayes Information Criterion (BIC) indices, based on the information theory.

Convergent validity

To examine whether the observed items of each dimension were strongly correlated between themselves, we estimated the AVE and the composite reliability (CR) (17, 20). According to Hair *et al.* (21), AVE values >0·5 and CC greater or equal to 0·7 indicates an adequate convergent validity.

Discriminant validity

Discriminant validity assesses whether the items that reflect one dimension are not correlated with another dimension (17). According to Fornell and Larcker (20) and Maroco (17), there is discriminant validity between dimensions i and j if AVE_i and $AVE_j \geq \rho^2_{ij}$.

Internal consistency

Internal consistency was estimated with the standardised Cronbach's alpha coefficient (α) (22) for each dimension proposed in the questionnaire.

Reproducibility

To estimate the intra-rater reproducibility, 62 of the 249 participants were randomly chosen and evaluated in two moments 1 week apart. The reproducibility was estimated with the intra-class correlation coefficient (ICC), with a 95% confidence interval. Temporal stability of the subject's responses was assessed by test–retest reliability using the Pearson's correlation coefficient (r).

Ethical aspects

This study was approved by the Ethics Committee on Human Research of the University Center of Araraquara – UNESP.

Results

Following the face validation process, the CVR was calculated for each item. The CVR ranged from a minimum of 0·43 (it1 and it2) to a maximum of 1 in several items (see Table 1).

Note that all items had a CVR above the significant minimum and were, therefore, retained in the questionnaire.

The reliability of the answers given in the different moments was good for both dimension D1 (functional capacity) (ICC_{D1} = 0·895, 95% CI 0·832–0·935) and dimension D2 (feeding) (ICC_{D2} = 0·825, 95% CI 0·726–0·891). There was also an excellent temporal stability for both dimensions (R_{D1} = 0·896, 95% CI 0·834–0·936; R_{D2} = 0·826, 95% CI 0·726–0·891).

The descriptive statistics measures used to evaluate the sensitivity of psychometric items of the MFIQs items are found in Table 2.

For both dimensions, all items had values of skewness and kurtosis indicative of substantial deviations from the normal distribution, with the exception of item seven which proved to be leptokurtic. However,

*AMOS ® 18.0.; SPSS Inc., Chicago, IL, USA.

Table 1. Content validity ratio for the 17 items of the Mandibular Function Impairment Questionnaire. Araraquara, Brazil, 2010

Items	Essential	Useful but non-essential	Non-necessary	RVC*
D1: Functional capacity†				
It1. Social activities	15	1	–	0·43
It2. Speaking	15	1	–	0·43
It3. Taking a large bite	21	–	–	1·00
It4. Chewing hard food	21	–	–	1·00
It5. Chewing soft food	18	3	–	0·71
It6. Work and/or daily activities	15	2	–	0·43
It7. Drinking	15	3	–	0·43
It8. Laughing	20	1	–	0·90
It9. Chewing resistant food	21	–	–	1·00
It10. Yawning	18	2	1	0·71
It11. Kissing	18	3	–	0·71
D2: Feeding‡				
It12. A hard cookie	19	2	–	0·81
It13. Meat	21	–	–	1·00
It14. A raw carrot	21	–	–	1·00
It15. French bread	21	–	–	1·00
It16. Peanuts/almonds	19	2	–	0·81
It17. An apple	20	1	–	0·90

*Minimum significant value according to Laewshe (1975) (16), 0·42.
†Owing to the complaints about your jaw, how much difficulty to you have with.
‡Eating food includes taking a bite, chewing and swallowing. How much difficulty do you have with eating.

Table 2. Descriptive statistics used to evaluate the psychometric sensitivity of Mandibular Function Impairment Questionnaires items

Item	Mean	Median	Mode	Standard deviation	Kurtosis	Skewness	Min	Max
It1	0·52	0·00	0·00	0·84	2·58	1·70	0·00	4·00
It2	0·62	0·00	0·00	0·89	1·06	1·34	0·00	4·00
It3	1·64	2·00	1·00	1·16	−0·98	0·16	0·00	4·00
It4	1·75	2·00	1·00	1·24	−1·01	0·21	0·00	4·00
It5	0·39	0·00	0·00	0·73	3·59	1·95	0·00	4·00
It6	0·45	0·00	0·00	0·92	4·24	2·22	0·00	4·00
It7	0·18	0·00	0·00	0·59	21·67	4·28	0·00	4·00
It8	0·70	0·00	0·00	0·98	0·48	1·21	0·00	4·00
It9	1·69	2·00	1·00	1·24	−0·88	0·28	0·00	4·00
It10	1·22	1·00	0·00	1·16	−0·84	0·50	0·00	4·00
It11	0·53	0·00	0·00	0·93	4·37	2·11	0·00	4·00
It12	1·05	1·00	0·00	1·03	−0·44	0·67	0·00	4·00
It13	1·35	1·00	1·00	1·18	−0·61	0·54	0·00	4·00
It14	1·42	1·00	0·00	1·26	−0·88	0·49	0·00	4·00
It15	0·75	0·00	0·00	0·95	−0·03	1·03	0·00	3·00
It16	1·28	1·00	0·00	1·25	−0·90	0·58	0·00	4·00
It17	1·40	1·00	1·00	1·20	−0·49	0·60	0·00	4·00

most of the items presented values that, according to Maroco's (2010) literature review (17), indicate no sensitivity problems or significant non-normality (23). Fig. 1 gives the confirmatory factor analysis model for the MFIQ.

The standardised factor weights of items 1, 2, 6 and7, of the functional capacity dimension, were below the appropriate standardised factor weights [$\lambda \geq 0{\cdot}5$; see, e.g. (17)]. The overall model fit was poor according to standard indices for the goodness of model fit

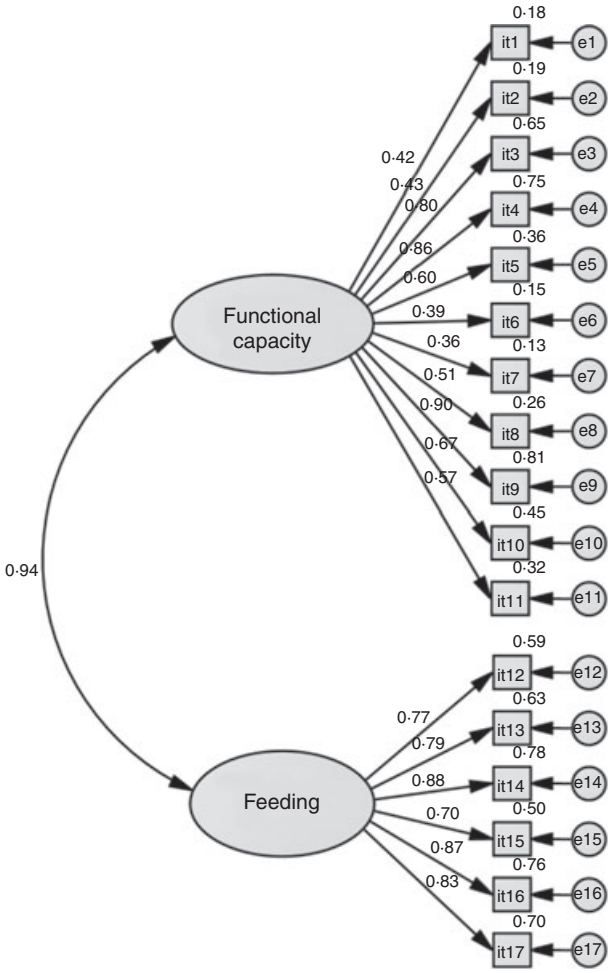

Fig. 1. Confirmatory factor analysis for the Mandibular Function Impairment Questionnaire with correlation between factors, factorial weights and R^2 for each item.

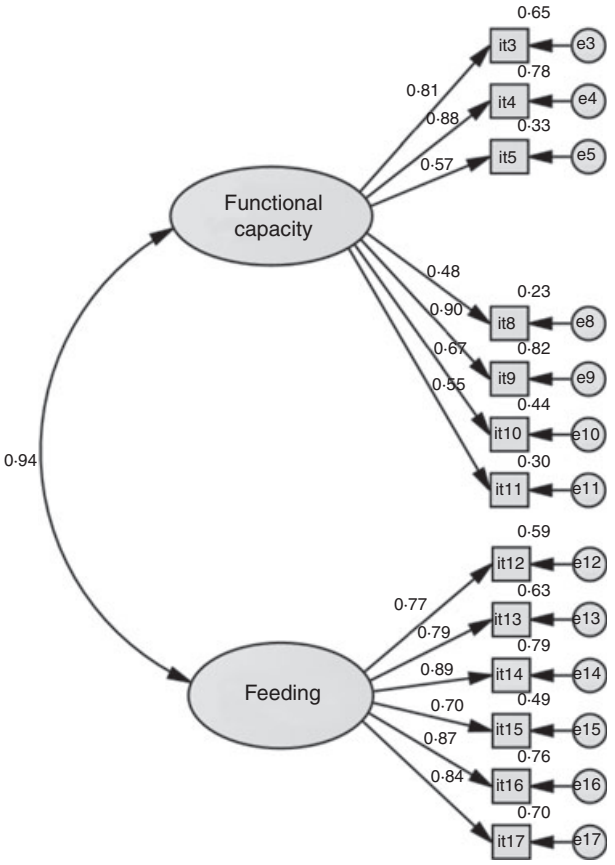

Fig. 2. Confirmatory factor analysis for the Mandibular Function Impairment Questionnaire after removal of problematic items (see text for details). Values are the correlation between factors, factorial weights and R^2 for each item.

(χ^2/df = 4·476; CFI = 0·850; GFI = 0·781; RMSEA = 0·118). Thus, we proceeded with the refinement of the original model. To improve the overall model fit and improve the factorial validity, items 1, 2, 6 and 7 were removed. The outcome of the refined factor model is presented in Fig. 2.

With the exception of item eight, all items of the MFIQ showed good standardised factor weights and adequate item reliability ($R^2 \geq 0{\cdot}25$). However, we decided to keep item eight on the scale, because its weight factor is very close to 0·50 and the modification indices did not suggest a high correlation with measurement errors and/or with dimension two (Feeding). The refined model has satisfactory goodness of fit indexes.

The two-factor model had a significantly better adjustment than the one-factor model ($\Delta\chi^2$ = 38 895, $P < 0.001$; two-factor: AIC = 274 678, BCC = 277 909, BIC = 369 649; one-factor: AIC = 311 573, BCC = 314 684, BIC = 403 027). The average variance extracted (AVE) and CR of the reduced MFIQ are given in Table 3.

The convergent-related validity was appropriate for both dimensions of the MFIQ ($AVE_j \geq 0{\cdot}5$; $CR_j \geq 0{,}7$). However, no discriminant-related validity was observed for the two dimensions of the MFIQ (AVE_{D1} e $AVE_{D2} < \rho^2$).

The internal consistency of the dimensions was excellent (α_{D1} = 0·874, α_{D2} = 0·918), and the high correlation between items of each dimension ($r_{\text{interitem\_D1}}$ = 0·462–0·543, $r_{\text{interitem\_D2}}$ = 0·624–0·687) indicated that they are part of the same conceptual construct. Thus, the MFIQ it is a consistent measurement scale.

Discussion

The use of scales in collecting information on health is customary. However, sometimes, little credibility is

Table 3. Average variance extracted and composite reliability (CR) of the reduced Mandibular Function Impairment Questionnaire

Dimension	AVE	CR	ρ^2
Functional capacity	0·507	0·872	0·705
Feeding	0·660	0·921	

AVE, average variance extracted.

given to this information because of the poor psychometric properties of data gathered with these measuring instruments. This distrust on the scales' results can be attributed to the fact these are usually made and/or used by health professionals who, by the nature of their training, may not be familiar with statistical methods (24–27).

It is essential to acknowledge that the measurement process is, to a greater degree, a process subjected to the probabilistic laws of nature so that, even believing that a result is accurate, it may not fully coincide with reality. Thus, the estimated level of accuracy of measured data is extremely important for assessing the quality of information collected. The quality evaluation of psychometric properties of the data gathered with measurement scales should be performed prior to its use.

In the process of validation of the MFIQ, the lowest content validity was observed in items 1 (social activities), 2 (talk), 6 (working and/or undertaking daily activities) and 7 (drinking) (see Table 1), which were also the items that had the lowest factor weights (Fig. 1). For items 1 and 6, this may have occurred owing to the high subjectivity of the terms 'social activities' and 'daily activities', because each evaluated individual performs different activities, which can hinder the understanding of what activities should be considered in each item. This interpretation was also reported by Ohrbach *et al.* (28). For items 2 and 7, the lower validity could be related to the fact that these are the functions that require less effort to be performed, when compared to the others presented in the questionnaire.

It should be noted that the presentation of descriptive statistics in Table 2 is justified by the necessity of validating the assumptions for performing confirmatory factor analysis and the subsequent evaluation of model's plausibility (17).

Figure 1 shows that the two-factor structure of the original MFIQ had poor quality of fit in this sample, with indexes below the suggested values for CFI, GFI and a high RMSEA that, according to the literature, should not be lower than 0·9 in the first two (17, 18) and >0·10 in the latter (29).

When items with low factor weights were excluded (Fig. 2), there was a satisfactory adjustment of the two-factor model, indicating that the reduced version of the instrument (MFIQ-r) has better factorial-related validity than the original version. The low discriminant validity observed (Table 3) was attributable to the high correlation (r = 0·94) found between the two dimensions. However, it should be emphasised that the absence of the model's discriminant validity must be reassessed in another independent sample, with similar characteristics to the one in this study, to assess whether this is a characteristic of the instrument or sample.

A comparison of MFIQ's validity in the present study and in the published literature is difficult, because only the study by Ohrbach *et al.* (28) gave information on the validity of this instrument. These authors performed an exploratory factorial analysis of MFIQ, applied to a sample of the US population, having found two dimensions (functional and social) that differ from the two factors proposed for Portuguese-speaking population. It is clear that the authors did not perform confirmatory factor analysis to assess the adjustment of the two-factor model in a different sample from the same population. Differences found between our study and that from Ohrbach *et al.* (28) are probably due to the fact that the psychometric quality of the data gathered with the same instruments is related to the studied population and, therefore, not an absolute characteristic of each scale (30, 31). It must be noticed also we did not established diagnosis groups, and we just identified the patients with some symptom or signal of TMD as indicated by RDC. The identification of diagnosis groups would be very useful for the establishment of criterion-related validity in future works with the MFIQ.

In this study, in addition to the evaluation of the psychometric characteristics of the data gathered with the MFIQ (sensibility, validity and reliability), we hope to set common grounds for procedural evaluation of measurement scales in odontology.

Conclusions

The MFIQ-r can be used to produce sensitive, valid and reliable data for the Portuguese-speaking population.

However, it must be noted that the two dimensions proposed did not show discriminant validity.

Acknowledgments

The authors thank FAPESP for financial support (process: 2009/11153-7) and Raquel Oliveira for proofreading the manuscript.

References

1. Okeson JP. *Dor orofacial*. Guia de avaliação, diagnóstico e tratamento. São Paulo: Quintessence; 1998.
2. Drangsholt M, LeResche L. Temporomandibular disorder pain. In: IK Crombie, PR Croft, SJ Linton, L LeResche, M VonKorff, eds. Epidemiology of pain – a report of task force on epidemiology of the International Association for the study of pain. Seattle: IASP Press; 1999:203–233.
3. Segu M, Lobbia S, Canale C, Collesano V. [Quality of life in patients with temporomandibular disorders]. Minerva Stomatol. 2003;52:279–287.
4. Oliveira AS. Evaluation of quality of life and pain in Temporomandibular Disorders (TMD). Brazilian J Oral Sci. 2005;4:646–650.
5. Tjakkes GH, Reinders JJ, Tenvergert EM, Stegenga B. TMD pain: the effect on health related quality of life and the influence of pain duration. Health Qual Life Outcomes. 2010;2:39–46.
6. Yap AU, Dworkin SF, Chua EK, List T, Tan KB, Tan HH. Prevalence of temporomandibular disorder subtypes, psychologic distress, and psychosocial dysfunction in Asian patients. J Orofac Pain. 2003;17:21–28.
7. Leao A, Sheiham A. Relation between clinical dental status and subjective impacts on daily living. J Dent Res. 1995;74:1408–1413.
8. Fonseca DM, Bonfante G, Valle AL, Freitas SFT. Diagnóstico pela anamnese da disfunção craniomandibular. Revista Gaucha de Odontologia. 1994;4:23–28.
9. Stegenga B, de Bont LG, de Leeuw R, Boering G. Assessment of mandibular function impairment associated with temporomandibular joint osteoarthrosis and internal derangement. J Orofac Pain. 1993;7:183–195.
10. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. J Craniomandib Disord. 1992;6:301–355.
11. De Laat A, Stappaerts K, Papy S. Counseling and physical therapy as treatment for myofascial pain of the masticatory system. J Orofac Pain. 2003;17:42–49.
12. Pereira LJ, Steenks MH, de Wijer A, Speksnijder CM, van der Bilt A. Masticatory function in subacute TMD patients before and after treatment. J Oral Rehabil. 2009;36:391–402.
13. 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. Fisioterapia e Pesquisa. 2008;15:121–127.
14. Guillemin F, Bombardier C, Beaton DE. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol. 1993;46:1417–1432.
15. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000;25:3186–3191.
16. Laewshe CH. A quantitative approach to content validity. Pers Psychol. 1975;28:563–575.
17. Maroco J. Análise de equações estruturais. Lisboa: Report-Number; 2010.
18. Byrne BM. Structural equation modeling with Amos: Basic concepts, applications and programming. New Jersey: Lawrence Erlbaum Associates; 2001.
19. McDonald RP, Ho MHR. Principles and practice in reporting structural equation analyses. Psychological Methods. 2002;7:64–82.
20. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Market Res. 1981;18:39–50.
21. Hair JF, Black WC, Babin B, Anderson RE, Tatham RL. Multivariate data analysis. 6th edn, Upper Saddle River, NJ: Prentice Hall; 2005.
22. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–334.
23. Kline RB. Principles and practice of structural equation modeling. New York (NY): The Guilford Press; 1998.
24. Emerson JD, Colditz GA. Use of statistical analysis in the New England Journal of Medicine. N Eng J Med. 1983;309:709–713.
25. Altman DG. The scandal of poor medical research. Br J Surg. 1994;308:283–284.
26. Schwartz SJ, Sturr M, Goldberg G. Statistical methods in rehabilitation literature: a survey of recent publications. Arch Phys Med Rehabil. 1996;77:497–500.
27. Kurichi JE, Sonnad SS. Statistical methods in the surgical literature. J Am Coll Surg. 2006;202:476–484.
28. Ohrbach R, Granger C, List T, Dworkin S. Preliminary development and validation of the jaw functional limitation scale. Community Dent Oral Epidemiol. 2008;36:228–236.
29. Ullman JB, Bentler PM, Schinka JA, Velicer WF. Structural equation modeling. Handbook of psychology: Research methods in psychology. Vol 2. Hoboken, NJ US: John Wiley & Sons Inc; 2003.
30. Honaker LM. The equivalency of computerized and conventional mmpi administration – a critical-review. Clin Psych Review. 1988;8:561–577.
31. Suris A, Borman PD, Lind L, Kashner TM. Aggression, impulsivity, and health functioning in a veteran population: equivalency and test-retest reliability of computerized and paper-and-pencil administrations. Comput Hum Behav. 2007;23:97–110.

Correspondence: João Maroco, Rua Jardim do Tabaco 34, 1149-041 Lisboa, Portugal. E-mail: joao.maroco@ispa.pt