The Jaw Functional Limitation Scale: Development, Reliability, and Validity of 8-Item and 20-Item Versions

Richard Ohrbach, DDS, PhD
Associate Professor
Department of Oral Diagnostic Sciences
University at Buffalo
Buffalo, New York

Pernilla Larsson, DDS, MDSc
Senior Consultant
Specialist Centre for Oral Rehabilitation
Linköping, Sweden

Thomas List, DDS, Odont Dr
Professor
Department of Stomatognathic Physiology
Malmö University
Malmö, Sweden

Correspondence to:
Richard Ohrbach
Department of Oral Diagnostic Sciences
University at Buffalo
Buffalo, NY 14214
Fax: 716-829-3554
E-mail: ohrbach@buffalo.edu

Aims: *To develop the Jaw Functional Limitation Scale (JFLS), comprising 3 constructs and a global scale, based on a preliminary instrument, and to investigate content validity of the overall functional limitation construct, reliability, and generalizability. A temporomandibular disorders (TMD) patient group, compared to other diagnostic groups, was hypothesized to report further limitation in each of the 3 new proposed constructs.* ***Methods:*** *One hundred thirty-two consecutive patients from 5 diagnostic groups (TMD, primary Sjögren syndrome, burning mouth syndrome, skeletal malocclusion, and healthy controls) participated in a known-groups validity design. Fifty-two jaw functional limitation items were identified by an expert panel for content validity. Rasch methodology was used for item reduction and assessment of model fit. The instrument was retested 1 to 2 weeks later.* ***Results:*** *Three constructs (mastication, vertical jaw mobility, and emotional and verbal expression) comprising a total of 20 items were identified along with a global scale (the JFLS-20), and each exhibited excellent psychometric properties with respect to modeled variance, item fit, reliability, and internal consistency. The psychometric properties of each construct remained satisfactory when analyzed separately among the 5 diagnostic groups. Temporal stability was satisfactory. A shorter 8-item form (JFLS-8) also proved useful for assessing global functional jaw limitation.* ***Conclusion:*** *The JFLS-20 is an organ-specific instrument comprising 3 constructs for assessing functional status of the masticatory system; the 3 scales exhibit properties that are ideal for both research and patient evaluation in patient groups with a range of functional limitations of the jaw. The JFLS-8 emerged as a short form for measuring global functional limitation of the jaw.* J OROFAC PAIN 2008;22:219–230.

Key words: functional limitation, masticatory system, oral health–related quality of life, reliability, validity

Orofacial function is an important contributor to an individual's general health and quality of life (QOL).[1] The masticatory system is responsible for complex biopsychosocial functions, where basal functions such as chewing, swallowing, eating, and yawning are manifested simultaneously with emotional functions such as smiling, laughing, screaming, and kissing. Oral diseases and disorders can lead to complications on 2 separate levels: limitations in orofacial functioning, limitations in psychosocial functioning and QOL, or both.[1–4] Three examples of orofacial functional limitation include: (*1*) impaired salivary function in xerostomic patients, which is associated with speech difficulties; (*2*) seriously resorbed jaws in denture wearers, which is associated with difficulty in chewing or smiling; and (*3*) temporomandibular joint pain, which is associated with limited jaw opening or difficulty with singing.

Functional limitation and disability are measured with generic, disease-specific, or organ-specific instruments. Generic instruments, such as the Short Form 36 (SF-36),[5] are used to evaluate the effects of illness or disease on different domains of overall functioning or health-related QOL, while disease-specific instruments focus on the effects of a single disease. In contrast, "organ-specific" instruments are generally focused on functional impact from the diseased organ or tissue, independent of the causative disease. One well-known instrument for the masticatory system, the Oral Health Impact Profile (OHIP),[2] is used to evaluate the effects of organ-related conditions or diseases on both jaw function and oral health–related QOL. Instruments such as the OHIP have the advantage of providing a comprehensive sweep that includes both the organ system and overall functioning, and the importance of this type of instrument lies not only in detecting limitations in condition-specific QOL but also in evaluating outcomes following interventions. The disadvantage is that measurement precision and identification of specific forms of limitation are, at least to some extent, sacrificed for breadth.[6]

At least 4 self-report scales have been developed to specifically measure at least some aspect of limitation in masticatory function.[7–10] Of these, the Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD) functional limitation checklist[7] and the Mandibular Functional Impairment Questionnaire (MFIQ)[8] appear to be the most frequently used. We previously evaluated the psychometric properties of these 2 instruments,[11] and identified problems included (*1*) inadequate definitions of the items concerning specific behaviors, (*2*) overlapping item content across both functional and psychosocial domains, and (*3*) validation performed only in TMD patients, which leads to unknown generalizability with respect to other orofacial conditions.

To address the problems identified with the existing instruments assessing jaw limitations, modern measurement methods[12] were used to develop a preliminary Jaw Functional Limitation Scale (JFLS)[11] as a global measure of functional limitation. This 8-item instrument exhibited excellent reliability, sensitivity to change, and validity, fulfilling the criteria for evaluative and discriminative instruments[13] and demonstrating what could be possible with good measurement methods. A major limitation, however, of this instrument was that item content and validation were based only on subjects with TMD, and broader assessment was deemed desirable.

Because previous instruments exhibited various limitations, our challenge was to develop a new instrument that overcame all of the identified limitations. The aims were to develop a JFLS comprising 3 constructs and a global scale, based on the preliminary 8-item instrument, and to investigate content validity of the overall functional limitation construct, reliability, and generalizability. A TMD patient group, compared to other diagnostic groups, was hypothesized to report additional limitation in each of the 3 new proposed constructs.

Materials and Methods

Subjects

Participants comprised patients (n = 134) recruited between August 2000 and June 2001 from the departments of oral medicine and dentofacial orthopedics at University Hospital, the Centre for Oral Rehabilitation, and the Public Dental Service, all in Linköping, Sweden. To ensure sufficient content validity across diverse oral conditions, subjects from each of 5 diagnostic groups were recruited: TMD (n = 31), primary Sjögren syndrome (PSS; n = 25), burning mouth syndrome (BMS; n = 20), skeletal malocclusion (n = 28), and healthy recall patients (n = 30). This type of study design is termed "known-groups validity," since known pathophysiology and clinical characteristics are used for interpreting measurement performance, but not for assessing whether the instrument can be used to distinguish individuals of one diagnostic group from another. We targeted a sample size of 30 subjects per group, with the expectation that we would retain at least 75 analytic subjects with valid data (ie, absence of extreme floor values or extreme ceiling values on all instrument items), which is considered adequate for 99% confidence that item calibrations will be stable within ± 1 logit.[14] Fifteen adult subjects with TMD from the University at Buffalo participated in qualitative interviews. The local ethics committees approved the study, and all individuals gave written, informed consent according to the Helsinki Declaration.

Selection Criteria

The TMD patients had received an RDC/TMD pain diagnosis.[7] The patients with PSS fulfilled the Copenhagen[15] and San Diego[16] classification criteria and the criteria proposed by the American-

European Consensus Group.[17] Patients in the BMS group reported pain and burning symptoms in the oral mucous membranes, with or without a diagnosis of oral lichen planus (either atrophic or erosive). The patients in the malocclusion group were selected from individuals who had been scheduled for orthognathic surgical correction of severe skeletal malocclusions; the most frequent malocclusions were anterior open bite and mandibular prognathism. The recall group comprised healthy dental patients recruited at routine recall at the same Public Dental Service clinic. Patients exhibiting more than 1 of the inclusion diagnoses were excluded from participation in the study.

Variables

A draft self-report instrument was constructed starting with the 8 items from the preliminary version of the JFLS,[11] and 44 additional items were recommended by a consensus panel of 5 expert clinicians and researchers in the dental fields of orofacial pain, oral medicine, and prosthodontics. The resultant 52-item instrument was the basis for item analysis and item reduction, leading to a final instrument of 20 items. The 52 items representing broad content validity and to be evaluated were:

- Mastication (20 items): chew gum, chew meat, chew chicken, chew French bread, chew hard bread, chew raw vegetables, chew raw apple, chew raw carrot, chew lettuce, chew nuts, chew crackers, chew potato chips, chewing (in general), bite into a sandwich, bite into a whole apple, eat hard foods, eat tough foods, eat soft foods, biting with the front teeth, biting with the back teeth
- Vertical jaw mobility (9 items): drink from a cup, drink through a straw, open wide enough to use a fork, open wide enough to brush teeth, open wide enough to bite into a sandwich, open wide enough to bite from a whole apple, open wide enough to talk, opening wide in general, keeping the mouth open
- Verbal and emotional expression (14 items): kiss, talk, pronounce words, whisper, sing, smile, laugh, frown, put on an angry face, put on a sad face, put on a happy face, whistle, yell, stick your tongue out
- Miscellaneous (9 items): sneeze, cough, lick, suck, swallow food, swallow liquid, swallow saliva, clear mouth of food, yawn

Overlap in item content among the 52 items was intentional, because the panel members had varying interpretations regarding how best to term the actual behaviors. Some items were tailored to reflect very specific daily activities, whereas other items (eg, "chewing [in general]") were intentionally broad (and hence, possibly vague) in case they proved to be valid omnibus descriptors of jaw functions. The degree of limitation, regardless of pain, in carrying out an activity was assessed on a numeric rating scale (NRS, 0 to 10 scale) anchored by the endpoints "no limitation" and "extreme limitation."

Design

After a new version of the instrument was drafted in English, the instrument was translated into Swedish by a researcher with experience in the subject field. The Swedish version was independently compared against the American English version by 3 researchers, and modifications in phrasing were then made simultaneously in both languages. Subsequently, a back-translation—from Swedish to American English—was made by a translator who had no knowledge of the initial instrument. The back-translation was compared against the original English, and corrections were then made in the American English version, the Swedish version, or both. The entire process was repeated until the researchers were confident that the 2 versions were semantically and culturally equivalent. This process is described in further detail elsewhere.[18]

Pilot testing and follow-up qualitative interviews were then conducted in Sweden with 8 subjects representing the diagnostic groups. Specific problems during item development occurred with items such as "eating celery," "drinking from water fountains," "Putting on a . . . face," and "frowning," all of which were deemed potentially useful by the expert panel. Adjustments and compromises were made in the final instrument to establish as broad a base as possible in assessing the domain of limitation. The final selection of good items would be empirically determined.

Patients were recruited consecutively from each clinic; all patients who met the selection criteria received written information explaining the research purpose and inviting them to participate. About 30 minutes were needed for each subject to complete the instrument battery, which contained other study instruments pertaining to the same research issues but are not presented here. If an item was unclear, an assistant was available to help. After 1 to 2 weeks, members of the TMD group completed the questionnaire a second time to assess test-retest reliability.

Qualitative interviews were conducted at the University at Buffalo with 15 TMD subjects. These subjects were administered the final 20-item instrument in 3 NRS response formats: 0 to 3, 0 to 5, and 0 to 10; after completion of all 3 instruments, the responses were normalized and compared across the 3 instruments for consistency and floor/ceiling effects, and subjects were interviewed regarding their preferred response format.

Statistical Analyses

Rasch methodology[19–21] served as the analysis method. It is a probabilistic model in which item difficulty and subject ability are estimated simultaneously. In linking item status to subject status via response frequency, a scale is produced that will demonstrate that fewer nondisabled people endorse more difficult items, relative to the directionality of the construct, while a greater number of more disabled people will endorse the same (more difficult) item. This analysis results in a scale onto which both items and subject are mapped commensurately. Rasch modeling yields parameter estimates for each item (and person), expressed in logits (log odds probability)[22] as well as model, item, and person fit.

The graded response model[23] of the Rasch method was selected for the 0-to-10 response format data. The category thresholds and item characteristic curves were assessed using the observed responses from each item, and the data were recoded from the 0-to-10 scale to a 0-to-3 scale to reduce statistical noise in the data. The recoded scores were used to develop a hierarchy of response likelihoods for each item in the set; the difficulty level of a particular item represents the ease, or likelihood, of endorsing that item as relevant to that individual. Item selection was based on mean-square residual values between 0.7 and 1.3 for infit (a measure of how individuals with limitations matching the difficulty of the item endorse that item) and outfit (a measure of how individuals with limitations not matching the difficulty of the item endorse that item). Any item that fit outside that range was considered for retention if its content was deemed important for the construct[22]; consequently, the scales emerged from the combination of content knowledge and statistical model fitting. Otherwise, items with lower fit values were considered redundant, and those with higher fit values were considered to fall away from the unidimensional construct.[24]

Overall model fit was assessed by a factor analysis on the residuals after Rasch extraction of items; this provided an estimate of how much variance the Rasch model accounted for in the data and was also used to determine whether any higher-order factors were present. If the observed pattern of responses deviated very little from the expected pattern and if there are no residual associations in the data, then items are said to represent unidimensional measurement and are considered to fit the Rasch model.[21,25] In addition to model fit, infit, and outfit, reliability coefficients can be computed for the items and persons. Item reliability is a measure of how stable the item measures would be with another group of similarly disabled subjects, while person reliability is a measure of how stable the subject disability estimates would be if a similar group of items was administered. Cronbach's alpha is a standard measure of internal consistency and augments the Rasch statistics.

Wright maps were constructed for inspection of distributions of item difficulties and person abilities but are not presented here. Temporal stability was assessed with the concordance correlation coefficient (CCC[26]). Stata (version 10.0; StataCorp) and Winsteps (version 3.57) were used for the analyses.

Results

Of the 134 patients recruited, 2 withdrew (1 from the TMD group and 1 from the PSS group). The TMD patient did not return the retest instrument within the stipulated time, and the PSS patient was unable to complete the study because of serious illness. The age distribution (means and ranges in years) in the different diagnostic groups was as follows: TMD (40.5; 19 to 64); PSS (59.5; 31 to 80); malocclusion (29.7; 18 to 59); BMS (60.1; 43 to 79); recall (46.0; 18 to 80). The entire subject group was 72% female.

General Inspection of the Data

Initial analyses assessed overall model fit and the performance of the 0-to-10 scaling. Substantial overlap in response probabilities with 0-to-10 scoring was noted, but distinct response probabilities emerged when the data were recoded to a 0-to-3 scale (0 = 0; 1 = 1 to 3; 2 = 4 to 7; 3 = 8 to 10). The item reliability remained between 0.94 and 0.96 regardless of how many response categories were permitted to remain in the data, whereas person reliability—the ability to distinguish persons—improved as the response categories were reduced from a 0-to-10 scale (reliability 0.67) to a 0-to-3 scale (reliability 0.82). Resultant item analyses were based on the 0-to-3 coding. The final items selected for each domain, to be described, are listed in Fig 3.

Table 1 Summary Statistics for Selected Masticatory Limitation Items by Diagnostic Group*

Group	Modeled variance (%)	Mean infit	Mean outfit	Item reliability	Person reliability	Cronbach's alpha
All groups	83.0	1.0	0.94	0.97	0.78	0.88
TMD	92.9	0.99	0.93	0.95	0.86	0.83
PSS	83.0	0.92	1.02	0.77	0.68	0.87
Malocclusion	94.0	0.95	0.90	0.93	0.78	0.89
BMS	69.0	0.98	0.93	0.68	0.67	0.87
Recall	52.4	0.94	0.85	0.36	—†	0.40

*Modeled variance refers to the amount of variance in the data explained by the Rasch dimension.Modeled variance refers to the amount of variance in the data explained by the Rasch dimension. Infit is ideal at 1.0 and signifies how respondents endorse items of similar difficulty. Outfit, also ideal at 1.0, signifies how respondents endorse items that are distant from the person's ability. Item reliability describes how other similar subjects would endorse these items, and person reliability describes how these subjects would endorse a similar set of items. Cronbach's alpha should be greater than 0.8 if the items are consistent with a unidimensional construct.
†Missing statistic is related to insufficient subjects in that group to estimate the parameter because of floor effects.

Mastication

All mastication items were entered into the model. The item values for many foods clustered together (ie, redundant information); therefore, items were removed through successive iterations considering fit and clinical meaningfulness, and a final set of items was determined based on suitable spread of item difficulty measures. The means of the remaining 5 items were only 1 standard deviation greater than the mean of the subjects, indicating reasonable and expected fit of the items to the persons where many persons were functioning relatively well. Ceiling effects were essentially absent in that only 2 subjects had limitation levels greater than the item with the highest limitation rating (eat soft food). Floor effects were judged by considering the easiest item (ie, tough food) to denote any reported limitation; "tough food" appears to be a reasonable threshold for capturing a minimal level of difficulty with mastication.

Table 1 presents summary statistical data for each diagnostic group. Note that infit and outfit were comparable for all 5 groups, and that modeled variance was equivalent for the TMD, PSS, and malocclusion groups, signifying generality of scaling across groups with differing pathophysiologies. The reliability measures differ across the groups in accordance with expectation based on the differing numbers of "healthy" subjects within the respective groups. For example, all but 6 of the controls (healthy recalls) endorsed only '0' responses to all mastication items because they did not have any difficulty in mastication as indexed by these test items; this affects reliability. In contrast, the TMD group had more subjects with nonextreme responses, and consequently, very high reliability statistics.

Vertical Jaw Mobility

All items were assessed in the same manner as described in the previous section. The items for this construct inherently fit less well than those for mastication, owing to far less overall limitation in mobility compared to mastication in this patient sample, but the item reliability itself was equivalent. The observed findings regarding model fit within each diagnosis (Table 2) seemed to meet expectations. Note that no statistics could be computed for the recall group because too many patients reported values at the floor of the instrument.

Verbal and Emotional Expression

All items were assessed in the same manner as described in the "Mastication" section, with the following exception in strategy: final item selection revolved around the tension between trying to balance efficiency (ie, item reduction) and maintaining suitable representation of the wide range of verbal and emotional behaviors for which the masticatory muscles (and associated facial muscles) are essential, so that the scale could also act as a checklist for monitoring the specific behaviors that an individual might regard as important. Consequently, there was evidence of notably poor outfit in the PSS group, whereas infit remained excellent for all groups (see Table 3). In terms of scale development, the wide range of oral behaviors used for verbal and emotional expression also exhibited uniformly high Cronbach's alpha values, suggesting that these items may be related more to a traditional factor; however, inspection of the residual factor plot did not indicate any significant clustering, and variance accounted for by Rasch scaling (73% to 93%) was on par with that of the items comprising mastication.

Table 2 Summary Statistics for Selected Vertical Mobility Limitation Items by Diagnostic Group

Group	Modeled variance (%)	Mean infit	Mean outfit	Item reliability	Person reliability	Cronbach's alpha
All groups	84.0	1.01	1.31	0.96	0.71	0.84
TMD	91.6	1.04	0.91	0.96	0.73	0.69
PSS	100	0.05	0.02	0.82	0.91	0.92
Malocclusion	61.1	0.87	0.98	0 00	0.56	0.84
BMS	97.9	0.01	0.01	0.00	0.82	0.97
Recall	—	—	—	—	—	—

See Table 1 for a detailed explanation.

Table 3 Summary Statistics for Selected Verbal and Emotional Expression Items by Diagnostic Group

Group	Modeled variance (%)	Mean infit	Mean outfit	Item reliability	Person reliability	Cronbach's alpha
All groups	76.1	0.99	1.01	0.86	0.77	0.92
TMD	76.7	0.97	0.91	0.81	0.71	0.83
PSS	84.6	0.75	1.63	0.53	0.82	0.95
Malocclusion	92.7	1.08	0.86	0.74	0.88	0.94
BMS	72.8	0.88	0.82	0.15	0.70	0.93
Recall	—	—	—	—	—	—

See Table 1 for a detailed explanation.

Other Constructs

The 3 items representing swallowing were not sufficiently distinct to form a unique unidimensional construct, the fit statistics were poor, and the construct could not be adequately defined operationally. The remaining items also did not fit any constructs. Swallowing as a general concept and yawning were retained, as described in the following.

Qualitative Interviews

Comparison of the normalized responses indicated that the 0-to-10 and 0-to-5 NRS response formats were roughly equivalent, while the 0-to-3 response format resulted in a greater percentage of floor effects. Subjects reported that they felt constrained in endorsing a low level of limitation when using the 0-to-3 response format and indicated a slight preference for the 0-to-10 NRS over the 0-to-5 NRS. The participants were polled regarding face validity of the selected items for each construct and their intended responses; the only problems identified related to the distinction between "eat" versus "chew" as it related to "soft food." TMD patients reported that when they experienced high levels of pain, they avoided the behavior termed "chewing," even with soft food, because any type of jaw movement was too painful; they simply ingested the food, moved it around with the tongue, and swallowed, and they reported that "eat" identified that behavior.

Global Functional Limitation

The preliminary JFLS contained 2 items—"chew" and "eat hard food"—that were not selected for the Mastication scale. The Rasch model of the original 8 items was rerun, and "chew chicken" and "chew tough food" were substituted. Item measures and fit statistics were compared; as expected per the prior analyses for the Mastication scale, "eat tough" and "eat hard" had similar item measure values (denoting roughly equivalent levels of difficulty assigned by the subjects to those behaviors), and "chew chicken" denoted less difficulty compared to "chew," filling a hole in the hierarchy, but showed equal fit otherwise. Consequently, a short instrument, the JFLS-8, emerged that could be administered separately; "chew chicken" and "chew tough food" were substituted for the respective original items. The resultant model statistics remained stable, and the correlation between total scores from the original 8-item instrument with the new 8-item global measure was 0.98. The summary statistics for global functional limitation (JFLS-8) are displayed in Table 4.

Discriminant and Construct Validity

The original 0-to-10 metric, per the qualitative interview outcomes of the selected items, was used to compute summary scores for each of the 3 scales. The summary scores were compared by analysis of variance for the different diagnostic groups. Due to

Table 4 Summary Statistics of Global Functional Limitation of the Jaw (Also Available Via JFLS-8) Items by Diagnostic Group

Group	Modeled variance (%)	Mean infit	Mean outfit	Item reliability	Person reliability	Cronbach's alpha
All groups	76.7	1.00	0.98	0.96	0.70	0.87
TMD	88.4	0.98	0.92	0.95	0.77	0.82
PSS	85.6	0.88	1.00	0.82	0.73	0.88
Malocclusion	75.7	1.08	0.96	0.83	0.65	0.84
BMS	92.0	0.98	0.84	0.59	0.84	0.93
Recall	23.5	—	—	—	—	—

See Table 1 for a detailed explanation.

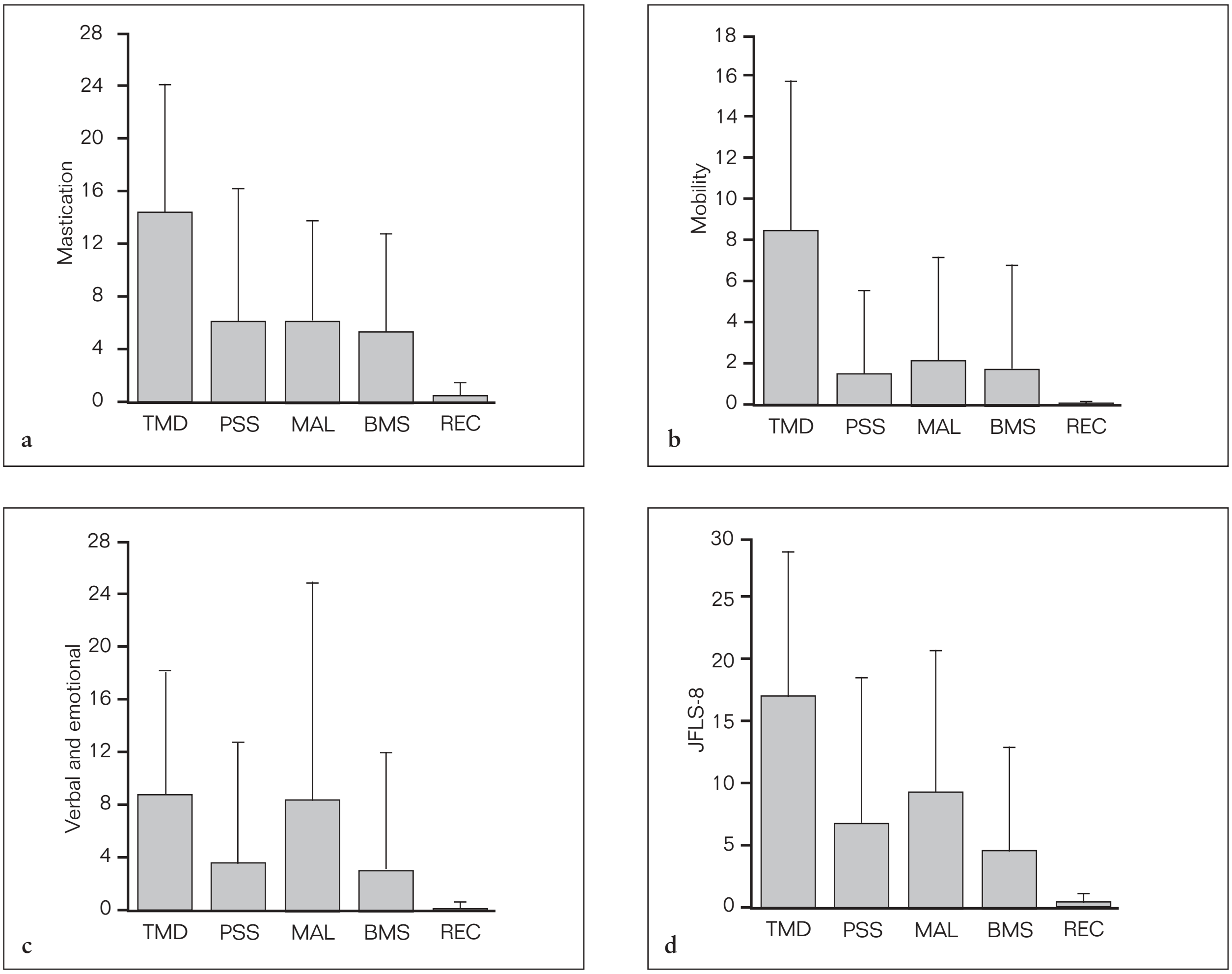

Fig 1 Comparison of limitation scores (means and SDs) for each construct according to diagnostic group (A, Mastication; B, Mobility; C, Verbal and Emotional; D, JFLS-8). These summary scale scores are based on the original 0-to-10 scale to demonstrate the generalizability of the findings, once the items were selected according to the Rasch model. Graphical depiction using a 0-to-3 scale had exactly the same appearance. Kruskal-Wallis was used for pair-wise comparisons of TMD versus each other group, using a Bonferroni-adjusted P value of .0125 based on 4 comparisons within each domain. For mastication, the TMD group reported significantly higher limitation compared to the other groups (all $P < .001$). For mobility, the TMD group reported significantly higher scores for reduced opening than the other groups (all $P < .001$). For verbal and emotional expression, the TMD group reported significantly greater limitation than the other groups (all $P < .003$) except for the malocclusion group ($P = .06$). For the JFLS-8 global limitation score, the TMD group reported significantly more limitation than the other groups (all $P < .005$). TMD = temporomandibular disorders; PSS = primary Sjögren syndrome; MAL = malocclusion; BMS = burning mouth syndrome; REC = dental recall patients.

Table 5 Correlations of Subscales

	JFLS-8	JFLS-20	Mastication	Mobility
All subjects (n = 131)				
JFLS-20	0.9675			
Mastication	0.8693	0.8653		
Mobility	0.8089	0.8620	0.7097	
Verbal/emotion	0.8357	0.8780	0.5641	0.6460
TMD subjects (n = 30)				
JFLS-20	0.9422			
Mastication	0.8836	0.8845		
Mobility	0.6819	0.8399	0.6622	
Verbal/emotion	0.8310	0.8467	0.5862	0.5649

Correlation matrix represents list-wise deleation of 1 subject who had missin data on 4 emotiono items, resulting in 1 missing score for that variable.

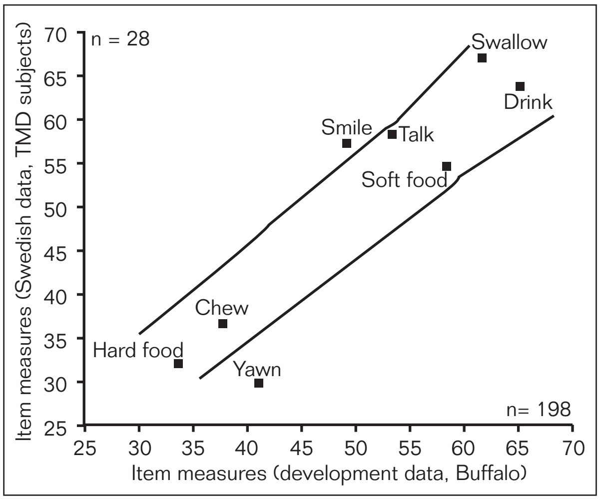

Fig 2 Item calibrations plotted per each site, with 95% confidence intervals computed on the basis of respective standard error estimates of each item measure. Transition thresholds within the 0-to-3 scale were calibrated using all subjects from the original Buffalo dataset (n = 198) and the Swedish dataset (n = 132). Item measure values were calibrated separately for each sample based on standardized thresholds, and for the Swedish group only the valid TMD subjects with nonextreme values were used (n = 28).

non-equality of variances (Kolmogorov-Smirnov test), the summary scores were compared by Kruskal-Wallis test for the different diagnostic groups, revealing significant between-group differences (each $P < .001$) in reported values of mastication, mobility, verbal and emotional expression, and global limitation (JFLS-8) and indicating that the scale scores and global functional limitation score are sensitive to behavioral limitations theoretically linked to the respective disorders. In accord with the hypothesis, TMD patients exhibited the greatest levels of limitation for each construct, and controls reported almost none. Figure 1 shows the means and standard deviations for each of these constructs. Analysis of variance using summary scores based on the 0-to-3 metric gave identical results. The relationships among the scale scores are shown in Table 5, demonstrating sufficient independence among the 3 constructs.

The investigators' previous publication[11] indicated that the preliminary 8-item JFLS had strong construct validity that was independent of depression, somatization, anxiety, and clinical findings. In addition to the construct validity of the respective scales as emerging from the Rasch model,[12] assessment of the construct validity of the 3 new scales was performed indirectly by (*1*) comparing the item performance of the preliminary 8-item JFLS in this subject sample against the item performance of the same items in the original Buffalo development sample, and (*2*) assessing, in this sample, the correlation between the 3 new constructs from the JFLS-20 with the preliminary 8-item JFLS. The data for the original 8 items (Swedish data) were combined with the data from the same items in the Buffalo development dataset[11]; category thresholds were determined in this combined dataset. The category thresholds were then used as anchors for separate calibrations of the items and, for this analysis, all 198 subjects in the Buffalo dataset were used, whereas only the 28 subjects with TMD, as a comparable diagnosis, were used in the Swedish dataset. The resultant measure scores and error estimates associated with each item were used to plot the measure values for each item within a 95% confidence interval (Fig 2). Considering that the Swedish dataset for the calibration analysis had only 28 subjects, the item measures appear to fit

For each of the items below, indicate the level of limitation during the past month. If the activity was completely avoided because it is too difficult, indicate '10.' If you avoid an activity for reasons other than pain or difficulty, then leave the item blank.

1*	Chew tough food	11*	Swallow
2	Chew hard bread	12*	Yawn
3*	Chew chicken (eg, prepared in oven)	13*	Talk
4	Chew crackers	14	Sing
5	Chew soft food (eg, macaroni, canned or soft fruits, cooked vegetables, fish)	15	Putting on a happy face
		16	Putting on an angry face
6*	Eat soft food requiring no chewing (eg, mashed potatoes, apple sauce, pudding, pureed food)	17	Frown
		18	Kiss
7	Open wide enough to bite from a whole apple	19*	Smile
8	Open wide enough to bite into a sandwich	20	Laugh
9	Open wide enough to talk		
10*	Open wide enough to drink from a cup		

Fig 3 Final JFLS. Items 1 to 6 represent mastication, items 7 to 10 represent mobility, and items 11 to 20 represent verbal and emotional communication. Items with an asterisk (*) are those used for the JFLS-8 (short form). Responses used a 0-to-10 NRS, with 0 anchored as "No limitation" and 10 anchored as "Severe limitation."

well. The correlations between the original 8-item global score and each of the 3 new scale scores were between 0.84 and 0.86, signifying high agreement between the new scales and the global score from the preliminary instrument. By way of contrast, the correlations between the global 20-item score and the 3 new scales were slightly higher, just under 0.90, signifying that the global score based on the JFLS-8 is quite robust with respect to its representation of the 3 component constructs. Internal consistency was 0.87 for the JFLS-8 and 0.95 for the JFLS-20.

Temporal Stability

The 52-item draft instrument was readministered to the TMD group 1 to 2 weeks later, and the total score at each administration was computed based on item selection, as described earlier. The CCC rho assessed temporal stability for each of the tested constructs as 0.87 for limitations in mastication and 0.94 for limitations in vertical jaw mobility. Temporal stability was 0.56 for verbal/emotional expression. The temporal stability for the 8-item scale was 0.81, whereas it was 0.87 for the full 20-item global score. The final items selected for the instrument are listed in Fig 3.

Discussion

A reliable and valid set of scales (JFLS-20) was developed using Rasch analysis, which measures the constructs of limitations in mastication, jaw mobility, and verbal and emotional expression; a global functional limitation score is also produced, and the latter can also be determined with a short form (JFLS-8). In addition to assisting in the development of concise scales, Rasch modeling allows each of the scales to be administered independently and allows the user to tailor instrument administration. This instrument would be suitable for use in research settings for better characterization of the functional status of patients and for use in clinical trials given the already established sensitivity to change.[11] For clinical settings, the internal consistency reliabilities of the mastication subscale and the JFLS-8 are borderline sufficient for use in assessing individuals, but the global limitation score from the JFLS-20 is quite sufficient for assessment of individuals.

Five diagnostic groups were selected to represent a range of masticatory system dysfunctions; the final items selected for each scale should be applicable across oral conditions and diseases that share clinical attributes with the selected samples. The known-groups validity design also permitted hypothesis testing as another form of validity assessment at this development stage of the instrument. In terms of other aspects of generalizability, the age and gender distributions in the different groups in this study correspond to distributions reported in other studies.[7,27,28] Severe malocclusions are usually treated during young adulthood after growth is finished; hence, the mean age of this group was substantially lower, as expected, than that of the other groups.[29] In sum, the present results are generalizable to the various orofacial pathologies.

The content validity of this instrument is primarily based upon item recommendations by specialists in oral medicine, pain, TMD, and prosthodontics. From this set of 52 items, a small number of items were empirically selected as sufficient to capture the measurement range within each of the constructs. This approach reflects item selection from the clinician's view, and since 1 goal of this study was to address content validity of the original JFLS,[11] an item-selection method that reflects the patient's point of view should be considered. Qualitative interviews or focus groups are sometimes used for item nomination as part of instrument development. A pilot test of the JFLS-20 in Sweden and qualitative interviews in the United States confirmed that the patients considered the final items to be understandable, sufficient, and clinically relevant.

The 3 constructs are conceptually distinct and empirically different from each other, as shown in Table 5. The most challenging construct is mastication, which comprises all movements related to preparing food for swallowing. Each type of food has its own consistency and texture, and a different level of chewing ability is required to form it into a bolus in preparation for swallowing. For example, tough food such as meat is difficult to chew for TMD patients, whereas individuals with an anterior open bite have difficulty biting off a piece of lettuce (which was represented by an item that did not pass the item fit test given the other diagnoses); nevertheless, these differences among disorders do not appear to require disorder-specific foods, as the model fit of the selected items within each diagnostic group was sufficient. Individuals who have a clinical disorder because of any of the included conditions reported escalating difficulty with eating foods ranging from hard to soft. Across disorders, the ability to master tough food would seem to represent an adequate measure of the health of the jaw for mastication.

Vertical jaw mobility includes all movements associated with opening the jaw, and the range defined here is "wide enough to drink from a cup" (which, if problematic, would seem to represent a severe limitation) to "wide enough to bite into a whole apple." As expected, the TMD group had the greatest difficulties in opening, consistent with relevant clinical data.[28,30,31] While it might be expected that Sjögren syndrome patients would report some difficulty with opening because of the absence of lubrication effects, the magnitude of that limitation has never been reported, to our knowledge. The malocclusion and BMS groups unexpectedly reported substantially greater limitations in opening compared to controls; possible interpretations include overreporting, instrument validity problems, or accurate reporting of a previously unknown aspect of these disorders. This is exactly the kind of question that a known-groups validity study design for instrument development is intended to address as part of the validity evaluation. The Rasch item statistics are very good for this measure when all subjects are used for that analysis, and conventional item-test correlations support each item as appropriate and the internal consistency statistic is reasonable for all diagnostic groups. Taken together, our interpretation is that this finding reflects a previously unknown aspect of these 2 disorders, one that is intriguing and certainly subject to further research.

Verbal and emotional expressions included awkwardly expressed but reliable items of "putting on an angry face" and "putting on a happy face." Electromyographic (EMG) activity measured during different emotional states indicates that negative emotions are associated with higher EMG activity of the corrugator muscles, whereas positive emotions are associated with increased activity of the zygomatic muscles.[32–34] While the consequences of orofacial conditions on overall functioning have been reported,[1–3] the consequences of orofacial conditions on emotional regulation specifically have received little attention. These descriptive data appear to elucidate the kinds of complaints that patients with TMD (and perhaps the other conditions included, especially the malocclusion group) have noted for years: the condition affects their entire life, and while the impact from pain on global function is obviously critical, it might be that 1 implication of the present findings is that individuals with orofacial conditions that result in limitations in emotional expression have lost a primary mechanism by which individuals can self-regulate.[35]

Of the items representing miscellaneous activities of the jaw, no coherent construct emerged. Only a few subjects responded to the items regarding swallowing difficulties, even though patients with PSS generally have this problem. As determined by the JFLS-8, problems with yawning and swallowing can represent key assessment points, and they are included in the final instrument for 2 reasons: the first is that they fit within the JFLS-8, which can be used alone if a scale shorter than the JFLS-20 is required, and the second is that they can represent important areas for either further clinical assessment or for monitoring important outcomes.

The qualitative interviews confirmed that the participants responded to the content of each item as intended. TMD patients indicated that when pain is severe, they no longer "chew" but simply ingest and swallow, a function labeled "eat" as opposed to "chew." Consequently, we added 1 more item, "chew soft food," which will bridge the gap between "eat soft food" and the next item in the hierarchy, "chew crackers."

As we demonstrated for the preliminary JFLS,[11] comparison of global jaw limitation scores to depression, somatization, anxiety, pain, jaw symptoms, pain-related disability, and palpation scores indicates that most correlate poorly with jaw limitation (most correlations < 0.3); the largest such correlations account for at most 30% to 35% of the variance in the limitation scores in TMD subjects. The 3 constructs of limitation in mastication, mobility, and communication likely retain that construct validation because of their very high correlation with the summary score of the preliminary JFLS. As a result, the present constructs form "clean" measures and are superior at least to the instruments that were examined previously, the RDC/TMD checklist and the MFIQ.[11]

The distinction between instruments that are focused on precise measurements of organ-specific functions (eg, JFLS) versus those that more broadly measure levels of limitation and disability (eg, OHIP) is important in terms of selecting the best instrument for a particular application. Overall, good validity was reported for the OHIP as a broader assessment instrument as compared to the SF-36 and General Oral Health Assessment Index (GOHAI),[36,37] but lower validity was found for some domains such as mastication. The shorter OHIP-14 places even less emphasis on mastication compared to behavioral and psychosocial outcomes, while the GOHAI focuses more on dysfunctions such as xerostomia and pain. Consistent with a recent report,[6] Larsson et al[38] found that the domain "functional limitation" in the OHIP was too limited and the items were not representative of limited jaw function. This is exemplified by such items from the OHIP as denture fit, affected appearance, ability to taste, and food impaction. The Manchester scale[10] occupies an interesting niche among these measurement instruments; 5 of the 7 items within its "physical disability" scale appear to map collectively to the 3 new constructs identified in this research, and the other 2 items in the Manchester scale assess neurosensory functioning, which was not considered a behavior, per se, relevant to assessing functional limitation in the more classic sense. That interpretation, however, of "functional limitation" is also subject to examination, as the World Health Organization hierarchy of disability concepts has much overlap from the physical to the psychosocial.[39]

The final JFLS instrument does have some limitations. While we have established sensitivity to change in the preliminary JFLS-8,[11] this factor remains to be determined for the 3 new scales; however, because they are even more focused within a hierarchy of items, it would be surprising if the attributes of the Rasch model to these scales did not carry forward. Additionally, while the translation and back-translation process resulted in items that exhibit a quality that transcends a single language—that is, "biting into an apple" now denotes a function rather than a simple linguistic concept—data to support cross-cultural validity are needed. In addition, the overall utility of the JFLS via use in studies of large treatment groups has not yet been established. The JFLS should also be tested in other patient groups that may reflect alternative endpoints in functional limitation, such as those with prosthodontic problems. Additional items may consequently be needed; because of the structure of the measurement scales as determined by Rasch methods, the JFLS easily accommodates additions to a scale, and users could easily tailor a modification and develop commensurate scaling within the Rasch logit measures. We have established that the functional limitation construct is distinct from pain-related disability,[11] and we will present data describing its relationship to disability per se (unpublished data), since that distinction in measurement between the constructs is critical. Finally, the construct pertaining to verbal and emotional expression requires further exploration within and across cultures so that clinicians may better understand the manner in which different orofacial limitations are expressed at the interpersonal level.

In sum, organ-specific functional limitation of the masticatory system comprises at least 3 separate constructs, and we propose that the JFLS-20 is a reliable and valid instrument for measuring with sufficient precision alterations in jaw functions that individuals with orofacial disorders report as significant. These measurements are distinct from competing constructs that are often found to a measurable extent in individuals with these disorders.

Conclusions

A new organ-specific instrument, the JFLS-20, was developed and found to exhibit good reliability and validity for 3 constructs assessing limitations in mastication, jaw mobility, and verbal and emotional expression, augmenting the use of the JFLS-8 for assessment of global functional limitation. The instrument is therefore recommended for both clinical and research use in patient groups with varying functional limitations of the jaw.

Acknowledgments

The authors thank the expert panel for their assistance in developing the content validity of the JFLS: Dr Jeffrey Burgess, Seattle, WA, USA; Dr Miriam Grushka, Toronto, Canada; and Dr Ales Obrez, University of Illinois at Chicago, USA. The authors thank Dr Inger Lundström, Department of Oral Medicine, and Docent Gunnar Paulin, Dentofacial Orthopedics, at the University Hospital in Linköping, Sweden, and Henrich Rotting, Public Dental Service in Linköping, Sweden, for their gracious assistance in this study. The completion of this manuscript was facilitated by the Visiting Professor Program at School of Dentistry, Malmö, Sweden, awarded to the first author. This study was funded by the Research Council of Public Dental Service, Östergötland County, Sweden.

References

1. Reisine ST, Fertig J, Weber J, et al. Impact of dental conditions on patients' quality of life. Community Dent Oral Epidemiol 1989;17:7–10.
2. Slade G, Spencer J. Development and evaluation of the Oral Health Impact Profile. Community Dent Health 1994;11:3–11.
3. Sugar D, Grushka M. Prevalence of oral and facial pain and discomfort: Preliminary results of a mail survey. Community Dent Oral Epidemiol 1987;15:169–172.
4. Osterweis M, Kleinman A, Mechanic D. Institute of Medicine, Committee on Pain, Disability, and Chronic Illness Behavior. Pain and Disability: Clinical, Behavioral, and Public Policy Perspectives. Washington, DC: National Academy Press, 1987.
5. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36) I: Conceptual framework and item selection. Med Care 1992;30:473–483.
6. John MT, Hujoel P, Miglioretti DL, et al. Dimensions of oral-health–related quality of life. J Dent Res 2004;83:956–960.
7. Dworkin SF, LeResche L. Research Diagnostic Criteria for Temporomandibular Disorders: Review, Criteria, Examinations and Specifications, Critique. J Craniomandib Disord Facial Oral Pain 1992;6:301–355.
8. Stegenga B, de Bont LGM, de Leeuw R, et al. Assessment of mandibular function impairment associated with temporomandibular joint osteoarthrosis and internal derangement. J Orofacial Pain 1993;7:183–195.
9. List T, Helkimo M. A scale for measuring the activities of daily living (ADL) of patients with craniomandibular disorders. Swed Dent J 1995;19:33–40.
10. Aggarwal VR, Lunt M, Zakrzewska JM, et al. Development and validation of the Manchester orofacial pain disability scale. Community Dent Oral Epidemiol 2005;33:141–149.
11. Ohrbach R, Granger CV, List T, et al. Pain-related functional limitation of the jaw: Preliminary development and validation of the jaw functional limitation scale. Community Dent Oral Epidemiol 2008;36:228–236.
12. Embretson SE, Reise SP. Item Response Theory for Psychologists. Mahwah, NJ: Lawrence Erlbaum, 2000.
13. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993;118:622–629.
14. Linacre JM. Sample size and item calibration stability. Rasch Measurement Transactions 1994;7:328.
15. Manthorpe R, Oxholm P, Prause JU, et al. The Copenhagen criteria for Sjögren's syndrome. Scand J Rheumatol 1986;suppl 61:19–21.
16. Fox RI, Robinson CA, Curd JG, et al. Sjögren's syndrome. Proposed criteria for classification. Arthritis Rheum 1986;29: 577–585.
17. Vitali C, Bombardieri S, Jonsson R, et al. Classification criteria for Sjögren's syndrome: A revised version of the European criteria proposed by the American-European Consensus Group. Ann Rheum Dis 2002;61:554–558.
18. Ohrbach R, Bjorner J, Jezewski MA, et al. Guidelines for Establishing Cultural Equivalency of Instruments. Available at URL www.rdc-tmdinternational.org. 2007.
19. Linacre JM. A User's Guide to Winstep's Rasch-Model Computer Programs. Chicago: Winsteps, 2004.
20. Ware JE Jr. Conceptualization and measurement of health-related quality of life: Comments on an evolving field. Arch Phys Med Rehabil 2003;84(suppl 2):S43–S51.
21. Andrich D. Rasch Models for Measurement. Newbury Park, CA: Sage Publications, 1988.
22. Wright BD, Masters GN. Rating Scale Analysis. Chicago: Mesa Press, 1982.
23. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978;43:561–573.
24. Smith E Jr. Effect of item redundancy on Rasch item and person estimates. J Applied Meas 2005;6:147–163.
25. Wright BD. Local Dependency, Correlations and Principal Components. Rasch Meas Trans 1996;10:509–511.
26. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45:255–268.
27. Ship JA, Grushka M, Lipton JA, et al. Burning mouth syndrome: An update. J Am Dent Assoc 1995;126:842–853.
28. List T, Dworkin SF. Comparing TMD diagnoses and clinical findings at Swedish and U.S. TMD centers using Research Diagnostic Criteria for Temporomandibular Disorders. J Orofac Pain 1996;10:240–253.
29. Epker BN, Wolford LM. Dentofacial Deformities: Surgical-Orthodontic Correction. St Louis: Mosby, 1980.
30. Dworkin SF, Huggins KH, LeResche L, et al. Epidemiology of signs and symptoms in temporomandibular disorders: Clinical signs in cases and controls. J Am Dent Assoc 1990;120:273–281.
31. Fricton JR, Schiffman EL. Reliability of a craniomandibular index. J Dent Res 1986;65:1139–1364.
32. Hu S, Wan H. Imagined events with specific emotional valence produce specific patterns of facial EMG activity. Percept Mot Skills 2003;97:1091–1099.
33. Lang PJ, Greenwald MK, Bradley MM, et al. Looking at pictures: Affective, facial, visceral, and behavioral reactions. Psychophysiology 1993;30:261–273.
34. Dimberg U. Facial electromyography and emotional reactions. Psychophysiology 1990;27:481–494.
35. Levenson RW, Ekman P, Friesen WV. Voluntary facial action generates emotion-specific autonomic nervous system activity. Psychophysiology 1990;27:363–384.
36. Allen P, McMillan AS, Walshaw D, et al. A comparison of the validity of generic- and disease-specific measures in the assessment of oral health-related quality of life. Comm Dent Oral Epidem 1999;27:344–352.
37. Locker D, Matear D, Stephens M, et al. Comparison of the GOHAI and OHIP-14 of the oral health-related quality of life measures of the elderly. Community Dent Oral Epidemiol 2001;29:373–381.
38. Larsson P, List T, Lundström I, et al. Reliability and validity of a Swedish version of the Oral Health Impact Profile (OHIP-S). Acta Odontol Scand 2004;62:147–152.
39. World Health Organization. International Classification of Functioning, Disability and Health (ICF). Geneva: World Health Organization, 2002.