Community Dent Oral Epidemiol 2008; 36: 228–236

COMMUNITY DENTISTRY AND ORAL EPIDEMIOLOGY

Preliminary development and validation of the Jaw Functional Limitation Scale*

Richard Ohrbach[1], Carl Granger[2], Thomas List[3] and Samuel Dworkin[4]

[1]Department of Oral Diagnostic Sciences, University at Buffalo, Buffalo, NY, USA, [2]Uniform Data Systems for Medical Rehabilitation, Buffalo, NY, USA, [3]Department of Stomatognathic Physiology, Malmö University, Malmö, Sweden, [4]Departments of Oral Medicine and Psychiatry, University of Washington, Seattle, WA, USA

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. © 2007 The Authors. Journal compilation © 2007 Blackwell Munksgaard

Abstract – *Objective:* Functional limitation is a construct distinct from disability, but the distinction between these constructs is often blurred when measuring jaw status because of content overlap in measurement instruments that include items pertaining to mechanical jaw functions, jaw functions that have obvious social implications, and only social functions. We describe analyses in support of assessing functional limitation as distinct from disability by using both factor analysis and the Rasch measurement model in reanalyzing two existing instruments and then create a preliminary version of a new instrument that has excellent psychometric properties. *Study Design and Setting:* Subjects in three separate studies from two university settings completed two existing self-report instruments (Research Diagnostic Criteria for Temporomandibular Disorders and Mandibular Functional Impairment Questionnaire). Rasch methodology, factor analysis, and the combination of both were used in these analyses. Unidimensionality, in support of an instrument assessing a single construct, was also assessed. *Results:* Problems identified in the existing instruments included scaling, internal reliability, length, unidimensionality, and content validity; the preliminary version of the Jaw Functional Limitation Scale (JFLS) resolves these problems critical to delineating functional limitation and it exhibits internal reliability coefficients of 0.82 for persons and 0.99 for items. Construct validity, via convergent and discriminant validity with other associated constructs, of the JFLS was established via low correlations with depression, anxiety, somatization, pain interference, pain-free opening, and palpation sensitivity, and via moderate correlations with pain and jaw symptoms. *Conclusion:* While areas of further development are outlined, we conclude that this preliminary version of the JFLS is a valid instrument for the measurement of temporomandibular disorder (TMD)-related functional limitation, and consequently delineates functional limitation from disability.

Key words: functional limitation; mastication; mobility; temporomandibular disorders

Richard Ohrbach, DDS, PhD, Department of Oral Diagnostic Sciences, University at Buffalo, 355 Squire, Buffalo, NY 14214, USA
Tel: +1 716 829 3590
Fax: +1 716 829 3554
e-mail: ohrbach@buffalo.edu

*This research has been presented at: International Association for Dental Research, 1999, 2002; International Association for the Study of Pain, 1999; and Uniform Data Systems for Medical Research, 1999.

Submitted 22 August 2006;
accepted 5 January 2007

One consequence of chronic pain is disability, and four key disablement constructs have been identified by various groups, including the World Health Organization (WHO) and the Institute of Medicine (1–3). While the terminology varies, the four constructs can be descriptively termed 'pathophysiology', 'impairment', 'functional limitation', and 'disability' (3). The first two constructs are objectively measured indices of disease status, while the next two are subjectively measured indices of disease impact, and both levels of assessment – objective and subjective – are considered equally valid and important to the assessment of disablement. Impairment refers to measured alteration in function caused by specific pathophysiology of a target organ system, functional limitation refers to impact at the organ level, and disability refers to impact at the personal level. While the distinction

doi: 10.1111/j.1600-0528.2007.00397.x

between pathophysiology and impairment is straightforward, the distinction between functional limitation and disability is often less clear, given the terminological variability and measurement overlap related to the latent trait-based item content of the respective constructs (1, 4). For these two constructs, clarity in how the distinction is conceptualized is interdependent with suitable discrimination between them (4).

In exploring the measurement of limitation or disability associated with jaw functioning, two approaches can be considered: a checklist from the Research Diagnostic Criteria for Temporomandibular Disorders (RDC/TMD) and the Mandibular Functional Impairment Questionnaire (MFIQ) (5, 6). The face validity of items in each instrument varies across both limitation and disability. The RDC/TMD checklist was developed by expert consensus, the response options are binary, no rules are available for scoring, and there are no psychometric data. The MFIQ was developed by a single group, response options are ordinal, and while factor analysis informed its structure, both limitation and disability appear as overlapping constructs.

The present paper will (a) examine the statistical properties of the RDC/TMD functional limitation checklist; (b) examine the properties of the MFIQ, and (c) provide preliminary data for the JFLS which adequately distinguishes functional limitation from disability. Statistical properties examined include reliability, validity and responsiveness.

Methods

Subjects

Individuals with TMD seeking care at university clinics were selected as the population of interest, and the data were obtained from three previously developed datasets: (a) Seattle Checklist dataset and MFIQ data; (b) Buffalo Checklist dataset; and (c) Buffalo Treatment Checklist, obtained from a pilot treatment intervention (7). All subjects either met formal TMD diagnostic criteria based on the RDC/TMD or had a specific TMD complaint but were subclinical to the RDC/TMD algorithm; frequencies and further description of the Seattle checklist dataset are available elsewhere (8). Pertinent descriptive statistics are presented in Table 1.

Instruments

Items for each instrument are listed in Tables 2, 4 and 6. The Seattle Checklist items used the original format (i.e. 'yes' versus 'no') in response to the functional limitation-oriented question, 'What activities does your present jaw problem prevent you from doing?' As part of programmatic research, the Buffalo Checklist items were linked to a disability instrument, the Pain Disability Index (9, 10), which used a numeric rating scale response format (0 = no disability; 10 = total disability) and alternative instructions 'Indicate the *overall* impact in your life by responding to each item'; the items were otherwise identical to the RDC/TMD limitation checklist. Note that two items (sexual activity, exercise) of the original RDC/TMD checklist were omitted from the Buffalo Checklist as they were *a priori* deemed to either be broader than jaw function (sexual activity) or to be more related to disability *per se* (exercise). The checklist items used in the pilot treatment study also used the same format as the Buffalo Checklist.

Table 1. Demographic characteristics of Buffalo and Seattle datasets

Demographic variables	Buffalo Checklist dataset	Seattle Checklist dataset	Buffalo Treatment study
n	219	242	30
Gender (% female)	83	83	87
Age			
Mean (years)	40.8	38.3	36.3
Range (years)	10–93	18–68	20–50
Chronicity			
Median (months)	18	48	Not available
Range (months)	1–399	1–480	Not available
≤3 months	10%	12%	Not available
Diagnoses – subjects meeting formal RDC/TMD criteria	90%	100%	100%

Table 2. Factor analysis of Seattle Checklist, based on tetrachoric correlation matrix and varimax rotation of factor loadings

Factor	1	2	3
Eigen value	6.8	1.6	1.3
Cumulative percent	57%	70%	81%
Chewing	0.90		
Drinking			0.95
Exercise		0.58	
Hard foods	0.87		
Soft foods	0.61		0.60
Smiling		0.51	0.56
Cleaning teeth	0.58		
Yawning	0.84		
Swallowing			0.70
Sexual activity	0.48	0.57	
Talking		0.66	
Usual facial appearance		0.96	

Factor loadings less than 0.45 have been omitted.

For present purposes, the MFIQ is an interesting instrument for two reasons: (i) item content within the first factor reveals items with complex face validity, leading to difficulties in interpretation; for example, 'taking a big bite' implies problems with mobility; 'chewing hard foods' implies masticatory limitation; and 'biting, chewing, swallowing when eating ...a raw carrot' implies limitation in both masticatory as well as accessory muscles; and (ii) while three factors were extracted, two were combined into one subscale making interpretation difficult ('nonmasticatory impairment' containing items such as 'yawning' and 'work and daily activities').

Other variables included depression and somatization from the RDC/TMD; anxiety from the Symptom Check List 90 (SCL90) (11); anxiety from the State-Trait Anxiety Inventory (12); characteristic pain intensity (average of current pain, and worst pain and average pain over prior 6 months) from the Graded Chronic Pain Scale (GCPS) (13); pain interference score from the GCPS; total number of symptoms from a TMD symptom checklist (14); and amount of vertical pain-free opening of the jaw and number of extraoral muscle sites tender to palpation, both from the RDC/TMD (5).

Analyses

The Seattle Checklist data are dichotomous and a tetrachoric factor analysis permits the assumption that the data presumably reflect an underlying continuous latent trait. The data in the MFIQ are ordinal, and conventional factor analyses are often used for such data; however, we used the polychoric approach in order to best estimate the presumed underlying latent trait as observed via the ordinal scale. A minimum eigenvalue of 1.0 defined a factor.

Rasch methodology, used for developing the revised instrument, is a probabilistic model in which item difficulty and person ability are simultaneously estimated. In linking item status to person status via response frequency, the underlying theory results in a unidimensional scale in which 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) items. If the observed pattern of responses does not deviate very much from the expected pattern and if there are no residual associations in the data, then items are said to follow the Rasch model (15), and the scale is considered to be a useful unidimensional measurement instrument (16). Rasch modeling yields parameter estimates for each item (and person) expressed in logits (log odds probability) (17), and item-fit and model-fit statistics are available.

The Rasch graded response model was selected for analysis of the Buffalo Checklist because of the 0–10 response scaling (18). Using the observed item responses, we developed a hierarchy of response likelihoods for each item in the set; the difficulty level of a particular item represents the ease, or likelihood, of the subject endorsing that item. Infit and outfit ranges of 0.7 to 1.3 were used for item selection; an item that fits outside of that target range may be retained if its content is deemed important for the construct (17). Model fit was assessed using infit and outfit item statistics and person and item separation statistics; together, these statistics represent how well the items conform to the assumptions of separability of item-response curves in the Rasch model. Unidimensionality is often claimed if the data are ordered hierarchically. To determine unidimensionality empirically, a factor analysis was performed on the residuals following item extraction in order to assess possible higher order factors in the data. Factor analysis was also performed on raw scores in order to test dimensionality in that manner.

Finally, in order to assess responsiveness to change and stability of item estimates, the resultant instrument was administered in a pilot treatment study; pre- and post-treatment scores were compared, and effect size for change and item difficulty estimates were computed.

The resultant JFLS, along with the Seattle Checklist and the MFIQ, was then compared with other relevant variables in order to judge the construct validity of the three instruments via convergent and discriminant validity using Pearson's correlations. Expected ranges for the validity correlation coefficients were determined *a priori* based on how functional limitation, as a distinct construct, was hypothesized to fit within the hierarchical conception of disablement. Mplus v3.11, Winsteps v3.57, and Stata v8.0 software were used for analyses.

Results

Evaluation of Seattle Checklist

Factor analysis of the Seattle Checklist data resulted in the emergence of three factors – F1, F2, F3 – using conventional criteria for defining factor items; see Table 2. Plausible interpretations are: F1 reflects problems with physically difficult functions in using the jaw; F2 reflects jaw functions with clear social connotations; and F3 reflects problems with physically easy functions in using the jaw. Important item loadings range from approximately 0.5 (minimally acceptable factor loading criterion for including an item in a factor) to 0.96, and factor structure is not clean. The eight items in the Seattle Checklist identical to those in the Buffalo Checklist were also subjected to Rasch analysis (Table 3). Of the 242 persons, 96 (39.7%) were dropped from the analysis because of their reporting only 'no' responses or only 'yes' responses, indicating that the measurement scale is insensitive to extreme levels.

Evaluation of MFIQ

Using a polychoric correlation matrix, two factors emerge, and items pertaining to masticatory-type movements or functions load onto F1 while items with more social connotations load onto F2; see Table 4. With the original MFIQ, the third factor items were arbitrarily rolled into F2; in this analysis, the two items from the third factor, yawn and kiss, naturally split across F1 and F2 consistent with other items from the respective functional versus social domains. While speaking, drinking, and laughing are all obviously jaw behaviors, it appears that the item context may facilitate the emergence of the latent trait 'social behavior' when subjects respond to the instrument.

Table 3. Rasch analysis of Seattle Checklist

Items	Rasch measure	Error	Infit MnSq	Outfit MnSq	Score correlation
Drink	94.8	4.6	0.83	0.28	0.50
Soft food	78.0	3.0	1.22	0.89	0.53
Swallow	77.7	2.9	1.01	1.03	0.54
Talk	53.8	2.3	0.97	1.06	0.70
Smile	51.0	2.3	0.82	1.26	0.73
Chew	22.2	2.7	0.82	1.05	0.73
Yawn	18.4	2.9	1.10	1.54	0.69
Hard food	4.2	3.7	0.98	0.55	0.68
Mean	50.0	3.2	0.97	0.96	
SD	30.5	0.7	0.13	0.37	

Note that item order, based on difficulty, has two reversals compared with the Rasch analysis of the Buffalo Checklist: swallow and soft, and chew and yawn.
Input: 242 Persons, 12 Items.
Analyzed: 146 persons, eight items, two categories.
Person: separation 0.90, reliability 0.45.
Item: separation 9.51, reliability 0.99.
MnSq, mean square.

Table 4. Factor analysis of Mandibular Functional Impairment Questionnaire, based on polychloric correlation matrix and varimax rotation of factor loadings

Factor	1	2
Eigenvalue	10.7	1.6
Cumulative percent	63%	72%
Social activities		0.88
Speak		0.68
Big bite	0.82	
Hard food	0.88	
Soft food	0.58	
Work/daily activities		0.75
Drink		0.61
Laugh		0.69
Tough food	0.86	
Yawn	0.65	
Kiss		0.52
(Eat) Hard cookie	0.80	
(Eat) Meat	0.77	
(Eat) Raw carrot	0.87	
(Eat) French bread	0.76	
(Eat) Peanuts	0.80	
(Eat) Apple	0.83	

Factor loadings less than 0.45 have been omitted.

Construction of the JFLS

Rasch analysis of the Buffalo Checklist displays the model statistics for the eight (of 10) items that met the inclusion criteria; see Table 5. Person separation and reliability are very good (analogous to Cronbach's alpha), and item separation and reliability are excellent for this eight-item scale. Prior analyses confirmed that of the original 10 items, two items, 'cleaning teeth' and 'usual facial appearance', were either redundant or did not fit the Rasch model and were removed. Because of substantial noise amongst the response frequencies, the 0–10 data were recoded into four categories (0 = 0, 1 = 1–3, 2 = 4–6, 3 = 7–10) with resultant clear evidence of unique responding at each of the four levels via item-response curves, and the Rasch analysis was rerun. The final range of item difficulties is appropriately distributed and errors are small. Although not shown, the item mean and person mean are almost equal, signifying matching of the items to this particular clinical population.

Of the 219 persons, 21 (9.6%) did not fit the model; 16 subjects reported '0' responses to all items while five subjects reported maximal scores to all items representing floor and ceiling effects, respectively. 'Floor effects' likely indicate that individuals who have a true '0' value, that is, no difficulty in limitation assuming that 'hard food' is a reasonable representative item for expected extreme jaw function during mastication. In con-

Table 5. Rasch analysis of Buffalo Checklist

Items	Rasch measure	Error	Infit MnSq	Outfit MnSq	Score correlation
Drink	66.9	1.4	1.32	1.48	0.64
Swallow	63.1	1.3	1.21	1.12	0.68
Soft food	59.2	1.2	1.24	1.15	0.68
Talk	53.6	1.1	1.09	1.05	0.73
Smile	49.1	1.1	0.92	0.94	0.77
Yawn	40.0	1.1	1.15	1.11	0.76
Chew	36.3	1.1	0.50	0.51	0.85
Hard food	31.7	1.1	0.71	0.69	0.82
Mean	50.0	1.2	1.02	1.01	
SD	12.1	0.1	0.27	0.28	

Input: 219 Persons, 10 Items.
Analyzed: 198 persons, eight items, four categories.
Person: separation 2.16, reliability 0.82.
Item: separation 9.62, reliability 0.99.
MnSq, mean square.

Table 6. Factor analysis of Buffalo Checklist

	10-item scale		Rasch eight-item scale
Eigen value	5.8	1.1	5.0
Cumulative percent	58%	69%	63%
Chewing	0.87		0.88
Drinking		0.69	0.62
Hard foods	0.86		0.80
Soft foods	0.45	0.57	0.72
Smiling		0.66	0.80
Cleaning teeth	0.52	0.46	–
Yawning	0.77		0.75
Swallowing		0.69	0.69
Talking		0.66	0.77
Usual facial appearance		0.66	–

The varimax rotated solution is presented for the two factor model.

trast, ceiling effects are of potential concern in terms of assessing the full extent of limitation associated with the disorder; fewer than 3% of subjects are so affected. Compared with the Seattle binary scaling, person separation is greater, signifying higher ability to discriminate among individuals with limitation; score correlations (similar to item-whole correlations) are higher; and error associated with estimated item difficulty is approximately three times smaller than for the binary response scale.

Factor analysis of residuals was performed (Winsteps), and the total explained variance based on extracting the principal Rasch dimension was 82.7% (38.4 of 46.4 eigenunits), leaving 17.3% because of other factors in the data. Examination of the factor plot for the first remaining factor, accounting for 4% of the variance, revealed two clusters of items: (i) chewing, eating hard foods, and yawning with loading values between 0.52 and 0.66, and (ii) smiling, drinking, and swallowing with values between −0.37 and −0.42. The next two remaining factors accounted for 3% and 2.4%, respectively. Collectively, the factor analyses of the residuals signify that the items in the model fit a unidimensional construct which can be called 'functional limitation' given the item content and nature of the hierarchy from most to least likely to be endorsed.

In order to fully compare the polytomous Buffalo Checklist version to the dichotomous Seattle Checklist version, a conventional factor analysis of raw scores from the Buffalo Checklist was performed as the scores are (pseudo) continuous; see Table 6. Note that even with the initial deletion of two items (sexual activity, exercise) from the original published instrument, there is a second weak factor; F1 suggests items pertaining to more physically difficult jaw function, while F2 suggests items pertaining to easier function. In contrast, removal of two items (cleaning teeth, having usual facial appearance) as indicated by the Rasch analysis results in a set of items with uniform item loadings on the factor analysis.

The Buffalo Checklist was analyzed after binary recoding (0 versus any); 23% of subjects were dropped from the analysis because of non-informative data, and person separation at 1.07 (reliability 0.53) and mean error at 2.3 are comparable with the Seattle checklist analysis, signifying that it is the dichotomous response option, not the subject sample, that leads to a less informative scale.

From the Buffalo Treatment Checklist data, pre- and post-treatment scores were obtained and subjected to a within-sample t-test. Responsiveness to treatment was demonstrated by pre-treatment mean of 24.0 (SD 12.95) and post-treatment mean 18.7 (SD 12.76). These values were associated with an effect size (mean change divided by baseline standard deviation) of 0.41; the difference in reported limitation using within-sample t-test was $t(24) = 2.06$ ($P = 0.0502$), signifying that the pilot study group reported less functional limitation as a result of short-term treatment. In order to demonstrate the stability of the item difficulty estimates, the pre- and post-treatment data were used for re-estimating item difficulties, and these are shown in Fig. 1, demonstrating that regardless of the change in reported severity of limitation of the items, the item difficulties are estimated similarly

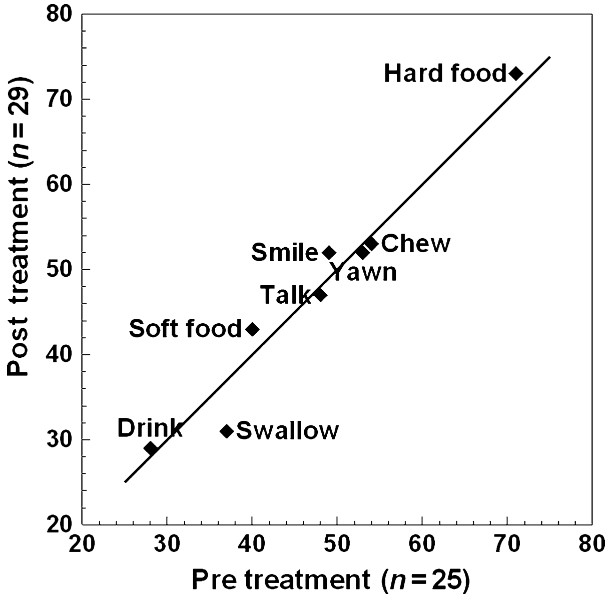

Fig. 1. Stability of JFLS item weights across treatment. Note that Rasch model handles missing data with respect to use of unequal sample sizes for pre- compared with post-treatment. Solid line is line of unity; item weights adhere to this line, which means that the effects of treatment do not affect the response probability leading to the item weights. Units are arbitrary Rasch weighting, range of 0–100, with default mean of 50 for the items.

on both occasions, indicating a reliable instrument across 2 months.

The three instruments, JFLS, Seattle Checklist, and MFIQ, were compared with other constructs in order to assess the construct validity of 'functional limitation'. *A priori* ranges of expected values were determined in order to help interpret the resultant correlation coefficients; these ranges stemmed from consideration of the WHO disablement model and taking the stance that correlations no greater than these expected values should be observed if subjects respond to the functional limitation construct without unexpected contamination from other constructs.

As displayed in Table 7, the psychological constructs of depression and anxiety exhibited a small relationship with the three instruments; somatization was expected to have a somewhat stronger relationship with functional limitation, because of the nature of the somatization items (i.e. awareness of bodily symptoms), and this was true for all three limitation instruments. Similarly, the JFLS exhibits a small relationship with pain interference (a form of disability). For the variables of pain and jaw symptoms, moderate correlations of approximately 0.5 are observed. Interestingly, the JFLS has a notably lower correlation pattern with clinical variables of opening range and number of palpation sites, compared with the other two instruments.

Discussion

The present study examined item content and statistical properties of two existing instruments that assess jaw limitation, and found that functional limitation can be better mirrored in a set of items identified and scaled via an appropriate measurement model. Using factor analysis and Rasch models, a parsimonious set of items coupled with an improved set of response options was developed and it exhibited substantially better psychometric properties for a scale that putatively measures functional limitation. Functional limitation, as a unidimensional construct, emerges from the item content and the directionality of scaling that is inherent in the terms. For items that exhibit a hierarchical relationship, the contrast between Rasch methodology and factor analysis highlighted

Table 7. Convergent and discriminant validity of the functional limitation instruments

		Correlation		
Predictor variable	Predicted relationship	JFLS	Seattle Checklist	MFIQ
MFIQ	None	n/a	0.74	n/a
Depression (SCL90)	Low	0.02	0.22	0.31
Anxiety (SCL90)	expected range:	n/a	0.30	0.29
Anxiety – State (STAI)	0.0–0.3	0.17	n/a	n/a
Somatization – full (SCL90)	Moderate	0.20	0.38	0.33
Somatization – no pain	expected range:	0.21	0.41	0.35
Pain interference (GCPS)	0.3–0.5	0.26	0.49	0.52
Characteristic pain (GCPS)		0.49	0.47	0.51
Jaw symptom index		0.57	0.68	0.62
Pain-free opening (mm)	Moderately high	−0.3	−0.47	−0.51
Extra-oral palpation (no. sites)	expected range: 0.5–0.7	0.12	0.41	0.36

the interdependency between the conceptualization of functional limitation and the selected items for the construct. The latter distinguished jaw-related functional limitation from person-level disability (i.e. pain-related interference) as well as from other relevant constructs such as mood, pain, and clinical examination findings.

Person reliability, item difficulty error, and average score correlation are better in the polytomous Buffalo Checklist versus the dichotomous Seattle Checklist. There are two sets of item reversals between these two versions of the RDC/TMD limitation checklist; each set contains an odd item (swallowing, yawning) which requires further investigation in order to determine where those items belong in the functional hierarchy. Nevertheless, the ranking of the remaining items retains the same order, regardless of whether the items were scored as 'yes'/'no' or as 0–10, indicating that this order is likely correct with respect to how the masticatory function becomes altered because of TMD conditions. This type of interpretation, a characteristic outcome of Rasch modeling, is critical for refinement of concepts such as functional limitation. Problems in eating hard food receives the largest number of endorsements while problems in drinking receives the smallest number of endorsements. Given that eating hard food is the most difficult jaw activity, the hierarchical order of limitation is necessarily the lowest when failure is reported only with eating hard foods. In contrast, limitation is highest when failure is reported in being able to drink; the latter is probabilistically linked to reported failure in all other functions as well. This is consistent with a common sense assessment of jaw functions and also consistent with the description by patients with the least jaw complaints (difficulty only with the most challenging functions) versus the patients with the most jaw problems (difficulty with even drinking). This link from clinical observation to hierarchical modeling also appears to support the functional limitation construct.

The complexity in defining functional limitation vis-à-vis disability is exemplified by two related observations. The first observation is item-based and pertains to content validity. The factor loadings in the analysis of the Seattle Checklist indicate that the responses for the item 'sexual activity' are related to both F1 (difficult jaw functions) and F2 (social aspects), and this result highlights the possibility of clarifying the boundary of two adjacent subjective disablement constructs by selecting items specific to limitation of the organ system and by selecting items with broader person-level disability connotations. Rasch item-fit statistics help delineate whether the selected item fits centrally within the targeted unidimensional construct or whether it belongs to a related construct (e.g. social aspects, disability). Improvement in assessing functional limitation – as a measurable and distinct construct – would lead, aside from scientific goals, to a better interpretation of the kinds of complaints that patients with TMD pain present with which, heretofore, have perhaps been unappreciated in the consulting room.

The second observation is factor-based and pertains to whether both factors of social context as well as mastication belong to the construct of 'limitation'. While social context could be construed as part of an instrument for assessing jaw limitation, summing the F1 and F2 scores in the MFIQ may lead to a highly misleading interpretation given the very different contingencies for performance problems in the two levels of biological organization (jaw system versus society). To that end, the delineation of hierarchical concepts by the WHO describing limitation in jaw function and disability, as two separable domains of assessment, seems more appropriate. In terms of respondent behavior, is a reported problem with 'drinking' due to difficulty in opening the jaw secondary to pain or is it related to social demands? The item can easily be construed to belong to both, but which response a subject provides will likely depend on context or content domain as determined by other items in the instrument (19). We assume that 'drinking', for example, may well belong to both domains: we would want to know about it, first, as a part of normal jaw functioning, and then, second, as a possible index of problems in more complex contexts. However, a measurement instrument would have to be designed to clearly delineate which level of meaning is intended by a complex item. Of special importance to one aim of this paper, the overall construct assessed by the published MFIQ is neither impairment (despite the title of the instrument and of the subscales) nor functional limitation, but more likely disability, yet with substantially less breadth as an oral health disability instrument than, for example, the Oral Health Index Profile (20).

The context of item responding in the JFLS appears to be different from the other two instruments, in that items respond differently via factor analysis and summary scores behave differently with respect to convergent and discriminant validity testing. It is noteworthy that the validity coefficients of the JFLS versus the other two instruments are lower for almost every pairwise assessment, suggesting greater independence in measuring the underlying latent trait, functional limitation. The improved psychometric performance of the JFLS is probably related to two characteristics: item selection and item-response scaling, both of which are properties of the Rasch method. The validity testing indicates that the functional limitation construct is independent from psychosocial dysfunction (via depression and somatization) as one aspect of disability. The moderate effect size of 0.41 is another index of validity, indicating sufficient sensitivity to change over a relatively short time period.

In terms of measurement models, there is a striking similarity in factor structure of items identified from each of three instruments, using different but item-response scale-appropriate correlation matrices for what is assumed to be a continuous underlying latent trait (i.e. limitation in using the jaw); this suggests that factor analysis is replicable when using the correct matrices. The interpretation of F3 and of F1 from the Seattle Checklist (Table 2), indicating physical jaw functions that are easy to do and that are hard to do, respectively, is probably facilitated by the latent-trait approach of Rasch which demonstrates the underlying hierarchy (4).

The present version of the JFLS instrument, while exhibiting very good psychometric properties, is preliminary. The items shown to be useful via Rasch analysis emerged from expert consensus opinion, but the full content domain should be assessed in light of the apparent hierarchical structure associated with the underlying latent trait. Items such as 'drinking' could use refinement with the possibility of improving both item properties and model fit. Finally, validity was shown only for individuals with TMD and not for a wider range of oral conditions.

In sum, the present study describes the transition of an instrument developed by consensus into one with psychometric validity consistent with conceptual validity for the disablement construct of functional limitation. The methods demonstrated in this paper highlight the utility of diverse psychometric approaches to instrument development. The JFLS exhibits good model fit as an instrument capable of providing continuous (or pseudo-continuous) measurement, it is valid with respect to very little overlap in assessment of related but different constructs, item-response characteristics are stable independent of change in the individual, and the scale is sensitive to clinical change. Further research is needed in order to assess adequacy of content validity, conventional test–retest reliability, and generalizability.

Acknowledgments

This research was graciously supported by the assistance of Uniform Data Systems, Buffalo, NY for data management and initial Rasch analyses. Consultation with Ben Wright, University of Chicago, is gratefully acknowledged. The collection of the data for the Seattle Checklist was supported by DE-10,766 from the National Institute of Dental and Craniofacial Research.

References

1. Granger CV, Seltzer GB, Fishbein CF. Primary care of the functionally disabled: assessment and management. Philadelphia, PA: JB Lippincott; 1987.
2. World Health Organization. International Classification of Functioning, Disability and Health (ICF). Geneva: World Health Organization; 2002.
3. Osterweis M, Kleinman A, Mechanic D. Pain and disability: clinical, behavioral, and public policy perspectives/Institute of Medicine, Committee on Pain, Disability, and Chronic Illness Behavior. Washington, D.C.: National Academy Press; 1987.
4. 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.
5. Dworkin SF, LeResche L. Research Diagnostic Criteria for Temporomandibular Disorders: Review, Criteria, Examinations and Specifications, Critique. Journal of Craniomandibular Disorders, Facial & Oral Pain 1992;6:301–55.
6. Stegenga B, de Bont LGM, de Leeuw R, Boering G. Assessment of mandibular function impairment associated with temporomandibular joint osteoarthrosis and internal derangement. J Orofac Pain 1993;7:183–95.
7. Uyanik J, Ohrbach R. Effects of relaxation on TMD symptoms: a randomized clinical trial. J Dent Res 1999;78:292.
8. Ohrbach R, Dworkin SF. Longitudinal changes in TMD: relationship of changes in pain to changes in clinical and psychological variables. Pain 1998;74: 315–26.
9. Tait RC, Chibnall JT, Krause S. The Pain Disability Index: psychometric properties. Pain 1990;40: 171–82.
10. Tait RC, Pollard CA, Margolis RB, Duckro PN, Krause SJ. The Pain Disability Index: psychometric and validity data. Arch Phys Med Rehabil 1987;68:438–41.

11. Derogatis LR, Lipman RS, Covi L. SCL-90: an outpatient psychiatric rating scale – preliminary report. Psychopharmacology 1973;9:13–28.
12. Spielberger CD, Sydeman SJ, Owen AE, Marsh BJ. Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI). In: Maruish ME editor. The use of psychological testing for treatment planning and outcomes assessment. 2nd edn. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers; 1999. p. 993–1021.
13. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992;50: 133–49.
14. Gerstner GE, Clark GT, Goulet J-P. Validity of a brief questionnaire in screening asymptomatic subjects from subjects with tension-type headaches or temporomandibular disorders. Community Dent Oral Epidemiol 1994;22:235–42.
15. Andrich D. Rasch models for measurement. Newbury Park, CA: Sage Publications; 1988.
16. Wright BD. Local dependency, correlations and principal components. Rasch Measure Trans 1996;10:509–11.
17. Wright BD, Masters GN. Rating scale analysis. Chicago, IL: Mesa Press; 1982.
18. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978;43:561–73.
19. Knowles ES. Item context effects on personality scales: measuring changes the measure. J Pers Soc Psychol 1988;55:312–20.
20. Slade G, Spencer J. Development and evaluation of the Oral Health Impact Profile. Community Dent Health 1994;11:3–11.