Community Dent Oral Epidemiol 2008; 36: 532–541

COMMUNITY DENTISTRY AND ORAL EPIDEMIOLOGY

Is the oral health impact profile measuring up? Investigating the scale's construct validity using structural equation modelling

S. R. Baker[1], B. Gibson[1] and D. Locker[2]

[1]Department of Oral Health and Development, School of Clinical Dentistry, University of Sheffield, UK, [2]Faculty of Dentistry, University of Toronto, ON, Canada

Baker SR, Gibson B, Locker D. Is the oral health impact profile measuring up? Investigating the scale's construct validity using structural equation modelling. Community Dent Oral Epidemiol 2008; 36: 532–541. © 2008 The Authors. Journal compilation © 2008 Blackwell Munksgaard

Abstract – *Objectives:* The aim of the study was to provide an empirical test of the construct validity of the Oral Health Impact Profile as a measure of Locker's conceptual model of oral health. *Methods:* A secondary analysis of data from the Ontario Study of Older Adults was carried out using structural equation modelling to assess the degree to which scale items measured the construct they were supposed to measure (*within*-construct validity) and whether relations between constructs were as hypothesized by Locker's model (*between*-construct validity). *Results:* The findings indicated that the Oral Health Impact Profile as currently conceived does not have adequate within-construct validity. Scale items did not always measure the construct they were supposed to measure, some items within a construct were redundant, many measured more than one construct, and the scale did not represent seven separate constructs of oral health as originally devised. Following reconceptualization of the scale, the revised six-factor 22 item version was a better fit to the data. However, the scale did not have adequate between-construct validity. *Conclusion:* The present findings do not provide support for the conceptual basis of the Oral Health Impact Profile as a measure of Locker's model of oral health. The need for further conceptual development of the scale, and Locker's model, are discussed.

Key words: confirmatory factor analysis; construct validity; locker; OHIP; oral health quality of life

Sarah Baker, Department of Oral Health and Development, School of Clinical Dentistry, University of Sheffield, Claremont Crescent, Sheffield, S10 2TN, UK
Tel: +44 114 271 7837
Fax: +44 114 271 7843
e-mail: s.r.baker@sheffield.ac.uk

Submitted 24 April 2007;
accepted 18 March 2008

Oral health-related quality of life (OHQoL) has become an important focus for assessing the impact of a range of oral conditions on individual's quality of life and well-being (1), together with the outcomes of clinical care such as, the effectiveness of treatment interventions (2, 3). There have been a number of instruments developed which supposedly measure OHQoL, with arguably the most popular being the Oral Health Impact Profile (OHIP) (4). Over the last decade, the original OHIP49, together with the shorter 14-item version (5), have become commonly used measures in both clinical and research settings.

One of the strengths of the OHIP is that it is purported to be derived from a theoretical framework; namely, Locker's (6) conceptual model of oral health. Published nearly two decades ago, Locker's model represented a fundamental shift in dentistry from a paradigm emphasizing disease and a medical model of care to one that incorporated a patient-centred perspective. As such, this multidimensional model provided a scientific framework for understanding oral disease and its clinical, as well as psychosocial consequences. The model states that there are five consequences of oral disease; impairment, functional limitation,

doi: 10.1111/j.1600-0528.2008.00440.x

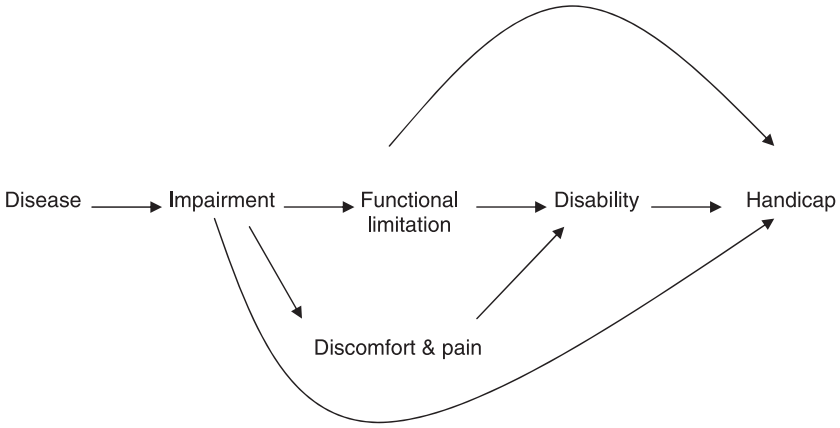

Fig. 1. Locker's (1988) conceptual model of oral health.

pain and discomfort, disability and handicap and that these are sequentially related (see Fig. 1). Impairment (structural abnormality e.g. edentulousness) leads to functional limitation (restrictions in body functions e.g. difficulty chewing) and pain/discomfort (self-reports of physical and psychological symptoms e.g. painful gums) which, in turn lead to disability (limitations in performing daily activities e.g. unsatisfactory diet) and then to handicap (social disadvantage e.g. social isolation). Functional limitation may also lead directly to handicap.

Original development of the OHIP involved interviews with 64 dental patients, which focused on their experience of dental disorders (4). The 535 statements which emerged from these interviews were analysed for their content, grouped into common themes, and matched *post hoc* to descriptions contained in Locker's conceptual model. From these, 46 were chosen by the authors to represent the complete series of statements. These, together with an additional three items measuring handicap from an existing inventory, were then administered to a group of 122 older adults (60+ years) to examine reliability and convergent validity.

Over the last decade, there have been numerous additional reports of the reliability and validity of both the OHIP49 and OHIP14 for a range of clinical groups, populations, ages, and countries. Despite this, rather surprisingly, there has been no testing of the conceptual basis of the OHIP. There are two questions of interest; first, the degree to which chosen items (e.g. sore jaw) measure the construct they are supposed to measure (e.g. pain). Secondly, whether relations between constructs are as hypothesized by Locker's underlying theoretical model (e.g. disability leads to handicap). Both questions relate to 'construct validity'; the first to *within*-construct relations, and the second to *between*-construct relations (7). For example, Items 34 'been upset' and 36 'felt depression' are said to represent the construct 'psychological disability'; yet, they could equally be indicators of another construct, 'psychological discomfort'. Similarly, what distinguishes Item 4 'appearance has been affected' (functional limitation) from Item 22 'felt uncomfortable about appearance' (psychological discomfort)? Are 'psychological disability' and 'psychological discomfort' conceptually distinct constructs as represented by the OHIP? At the present time, the premise that the OHIP49 and OHIP14 measure seven oral health-related constructs and that they represent a measure of Locker's conceptual model of oral health is empirically untested.

There are a number of clinical and conceptual reasons why providing answers to such questions are important. First, when an instrument says that it measures something (e.g. OHQoL or Locker's model of oral health), clinicians using that measure will assume that it measures what it says it measures. Without adequate construct validation, that assumption is not justified. Second, in studies utilizing the OHIP49 or OHIP14, authors often report scores for each of the seven oral health constructs, and state whether a clinical condition (or treatment) had a differential impact across these domains (2). If a condition (e.g. xerostomia) is associated with greater functional limitations compared to social disability, the implication is that clinical interventions may be most effective if targeting biophysical consequences rather than social limitations. However, if the scale does not have adequate construct validity, such conclusions would not be valid. Thirdly, the validity of OHIP49 or OHIP14 as an outcome measure in clinical trials is partly dependent on understanding the causal

processes linking oral conditions to patient-reported outcomes. In order to understand the pathways underlying such effects, the relationships between constructs as hypothesized within Locker's (6) model need to be empirically tested. To date, only one study has attempted to explicitly test Locker's model of oral health using empirical evidence based on OHIP14 data (8). Finally, ongoing development of the OHQoL field, just as in any field of inquiry, requires key constructs to be explored and disentangled. Only by testing the empirical validity of the underlying conceptual model, is it possible to propose alternatives which address any identified weaknesses.

The aim of the present research was to provide an empirical test of the construct validity of the original OHIP49 scale as a measure of Locker's (6) conceptual model of oral health. The model was tested with data from the Ontario Study of Older Adults (9) using structural equation modelling. The three key research questions were:

1. Do items selected to measure a construct actually do so? (*within*-construct validity).
2. Does the scale measure seven separate constructs (functional limitation, pain, discomfort, social disability, psychological disability, physical disability, handicap)? (*within*-construct validity).
3. Are the relationships between the constructs as hypothesized within Locker's conceptual model? (*between*-construct validity).

Materials and methods

Data were collected as part of the first follow-up phase of the Ontario Study of Older Adults, an observational cohort study of individuals aged 50 years and over with a baseline phase and follow-ups at 3 and 7 years (9). Of the 541 subjects, 234 were male and 307 were female. They ranged in age from 53 to 89 with a mean of 62.6 years. Most, 83.5% were dentate. One quarter, 24.9%, rated their oral health as being only fair or poor. The study was approved by the University of Toronto's Ethics Research Office.

Measure

Oral health impact profile

The OHIP49 (4) assesses frequency of problems associated with the mouth or dentures on 7 dimensions: functional limitation (9 items; range 0–36), pain (9 items, 0–36), psychological discomfort (5 items, 0–20), physical disability (9 items, 0–36), psychological disability (6 items, 0–24), social disability (5 items, 0–20), and handicap (6 items, 0–24). Participants are asked to rate for the last 3 months each item on a 5-point scale from 1 (''never'') to 5 (''very often''). The three items relating to dentures were removed (Q9 'dentures not fitting', Q18 'discomfort due to dentures', Q30 'unable to eat because of dentures') as each item had 50% missing responses. Means, SDs and ranges for the total scale and seven subscales are shown in Table 1[1].

Table 1. Mean (SDs) and sample ranges

	Mean	SD	Sample range
Functional limitation	15.11	5.19	8–36.00
Pain	14.60	5.44	8–36.85
Discomfort	8.35	3.97	5–25.00
Physical disability	11.37	4.90	8–40.00
Psychological disability	8.18	3.89	6–30.00
Social disability	5.94	2.82	5–25.48
Handicap	7.67	3.33	6–30.30
Total OHIP	71.22	24.13	46–208.08
N	541		

Table statistics are based on 46 items only (Items 17, 18 and 30 relating to dentures were removed).

Structural equation modelling

The first two research questions were examined using confirmatory factor analysis (CFA). CFA is the first in the two stage process of structural equation modelling (SEM) (the measurement model) (11). CFA has many advantages over traditional factor analysis. Traditional factor analysis is an exclusively inductive or data-driven technique with no prior specification of factors; in contrast, CFA is theory-driven and is used when a number of underlying factors are hypothesized *a priori*. CFA provides information on how scale items (e.g. ''have you had trouble pronouncing any words?'') measure underlying (latent) constructs (e.g. functional limitation). Given that latent constructs are not directly observable rather they have to be inferred from responses to items said to measure that construct, CFA provides a test of the validity of the selected items (Research Question 1). In addition, CFA provides information about the number

[1]Given some missing values, the data were imputed prior to analysis (10). We used regression imputation techniques which predict the unobserved values for each case as a linear combination of the observed values for that case. These predicted values are used to replace the missing values.

of constructs which ''fit'' the data (e.g. 1 ('oral health impacts') or 7 (e.g. 'pain', 'discomfort' etc) (Research Question 2).

The initial step of the analysis was to employ CFA to test the hypothesized measurement model specified in the original paper by Slade and Spencer (4). This involved a first-order CFA with functional limitation, pain, discomfort, physical disability, psychological disability, social disability and handicap as the seven latent constructs. Scale items (indicators) representing each of the seven constructs were as detailed in Appendix 1 of Slade and Spencer (4). Items were not allowed to load on more than one construct, nor were their error terms allowed to correlate.

The measurement model was examined using AMOS 7.0 with maximum likelihood estimation and bootstrapping (10). The bootstrap framework has been suggested as one approach for SEM when data may not meet assumptions of normality (11, 12). In our data, some of the items, most notably in the social disability and handicap subscales, were skewed. In bootstrapping, multiple samples (n = 1000 + , (13)) are randomly drawn from the original sample; the CFA model is then estimated in each dataset, and the results averaged. The bootstrap estimates and standard errors [together with bias-corrected 95% confidence intervals (CIs)] are then compared to the results from the original sample to examine the stability of parameters and test statistics (12, 13).[2]

As recommended, we evaluated model fit using a range of indices from the three fit classes; absolute fit, parsimony adjusted and comparative (12). The chi-square statistic (χ^2) was included as a measure of overall goodness of fit. However, given that χ^2 can be inflated by sample size, we also report the χ^2/d.f. ratio. The parsimony-adjusted index was the root-mean-squared error of approximation (RMSEA) with 90% confidence intervals (90% CI). The comparative fit indices were the Tucker-Lewis index (TLI) and the comparative fit index (CFI). A nonsignificant chi-square, a χ^2/d.f. ratio < 3.0, RMSEA values 0.08 or below, and CFI and TLI of 0.90 or above were taken to indicate an acceptable model fit (12, 14).

[2]There were no significant differences between maximum likelihood and bootstrap parameters in any of the models tested. Thus, all results reported here are bootstrapped estimates, standard errors and significance levels.

Results

Research Questions 1 and 2: Testing the within-construct validity of OHIP

To address the first two research questions, we tested a measurement model of the original seven-factor scale as detailed by Slade and Spencer (4). The CFA indicated that this model was not an acceptable fit to the data on any of the *a priori* model fitting criteria (see Table 2, Model 1). Given that the structure of the OHIP, as originally proposed, was not supported, the next step was to re-specify the scale so that it better represented the data. In line with recommendations (12), we examined the standardized residuals to investigate where relationships between items were being over- or under-estimated within the model. Those items with values > 2.58 (which corresponds to $P < 0.01$ level), were then deleted (n = 24). The 22 items retained in the model can be seen in Table 3.

The revised 22-item model was a significantly better fit to the data than the original model (see Table 2; Model 1 versus 2 comparison), and met four out of the five *a priori* model fitting criteria (Model 2, Table 2). Although, in general, a good fit to the data, factor correlations between some of the latent constructs were high suggesting that they may not represent distinct constructs. In CFA, factor correlations which exceed 0.85 are typically used as the criterion for poor discriminant validity (12). In our model, three correlations met this criteria; functional limitation-discomfort (r = 0.856), functional limitation-physical disability (0.856) and discomfort-psychological disability (0.925). We begun by collapsing the latent constructs of discomfort and psychological disability into a single factor (relabelled 'psychological impact') and re-estimating the model. The re-estimated model met three out of the five criteria (see Model 3, Table 2). The correlations between factors in the revised model indicated acceptable discriminant validity: functional limitation with pain (r = 0.714), psychological impact (0.716), social disability (0.484) and handicap (0.493); pain with psychological impact (0.705), physical disability (0.714), social disability (0.412) and handicap (0.617); psychological impact with physical disability (0.825), social disability (0.510) and handicap (0.723); physical disability with social disability (0.516) and handicap (0.646); social disability with handicap (0.800). The only exception was functional limitation and physical disability (r = 0.855), which indicated a borderline correlation. The

Table 2. Fit indices for the measurement and structural models

Model	χ^2 (d.f.)	d.f.	p	χ^2 /d.f.	RMSEA (90% CI)	CFI	TLI	Criteria fitted
1	5143.687	968	0.001	5.314	0.089 (0.087–0.092)	0.744	0.726	0
2	526.741	188	0.001	**2.802**	**0.058** (0.052–0.064)	**0.937**	**0.923**	4
3	599.904	194	0.001	3.092	**0.062** (0.057–0.068)	**0.925**	**0.910**	3
4	1309.250	209	0.001	6.264	0.099 (0.094–0.104)	0.796	0.774	0
5	657.997	198	0.001	3.323	**0.066** (0.060–0.071)	**0.915**	**0.900**	3

Model comparisons.
Model 1 versus 2 $\Delta\ \chi^2$ (780) = 4616.946, $P < 0.001$.
Model 3 versus 2 $\Delta\ \chi^2$ (6) = 73.163, $P < 0.001$.
Model 4 versus 3 $\Delta\ \chi^2$ (29) = 709.346, $P < 0.001$.

Model 1 = CFA seven-factor, 46 items; Model 2 = CFA seven-factor, 22 items; Model 3 = CFA six-factor, 22 items; Model 4 = CFA one-factor, 22 items; Model 5 = structural six-factor, 22 items. χ^2 = chi-square; d.f. = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root-mean-square error of approximation; CI = confidence interval. Figures in bold are those that meet the *a priori* model fitting criteria.

Table 3. Items retained in the reconceptualized OHIP scale following confirmatory factor analysis

Construct/factor	Item	
Functional limitation	1	Difficulty chewing
	2	Pronouncing words
	3	Tooth doesn't look right
	4	Appearance affected
	6	Taste worse
	7	Food catching
Physical pain	9	Painful aching
	11	Headaches
	16	Sore spots
Psychological discomfort	21	Miserable
	23	Tense
Physical disability	27	Unable to brush teeth
	28	Avoid eating
	29	Diet unsatisfactory
	32	Interrupt meals
Psychological disability	34	Upset
	35	Difficult to relax
	36	Depressed
Social disability	40	Less tolerant of others
	41	Trouble getting on with others
Handicap	45	Financial loss
	48	Unable to function

bootstrapped estimates for the six-factor 22 item CFA model can be seen in Fig. 2, alongside squared multiple correlations for the indicator items. As can be seen from Fig. 2, the beta weights for the six functional limitation indicator items ranged between 0.274 ('tooth does not look right'; $R^2 = 0.08$) and 0.822 ('chewing'; $R^2 = 0.67$). Similarly, for pain, the items ranged from 0.507 to 0.786, with the lowest loading item being 'headaches'. For psychological impact and social disability, there was less range in the item weights. For physical disability, the item with the lowest loading and smallest variance accounted for was 'unable to brush'; whilst for handicap, the 'financial loss' item had the lowest loading (0.457) and variance accounted for ($R^2 = 0.21$). Nevertheless, all 22 items were highly significant indicators of their respective constructs (all ps < 0.01).

Our final step was to examine whether this multidimensional six-factor model was a better fit to the data compared to an alternative simpler unidimensional model. To test this we examined a one-factor model in which all 22 items were loaded on to one latent construct of 'oral health impacts'. This model did not fit the data (see Table 2, Model 4), and was a significantly worse fit than the six-factor model (Model 4 versus 3 comparison, Table 2). These results confirm that oral health impacts, and the OHIP scale, are indeed multidimensional as originally conceived by Locker (6) and Slade and Spencer (4).

In summary, the within-construct validity of the original OHIP49 was not supported by the data presented here. Many of the individual scale items originally selected to measure a construct did not actually do so, rather they mapped onto multiple constructs and/or their error terms were highly correlated indicating that they measured similar aspects of oral health (Research Question 1). Following re-specification, a 22-item OHIP scale was found to fit the data better. This revised scale represented a six-factor underlying model rather than the seven separate constructs originally proposed by Slade and Spencer (4) (Research Question 2).

Research Question 3: Testing the between construct validity of the reconceptualized OHIP

To address Research Question 3, we tested a structural model of the revised 22 item six-factor scale. Using structural equation modelling, we examined the direct and indirect (mediated)

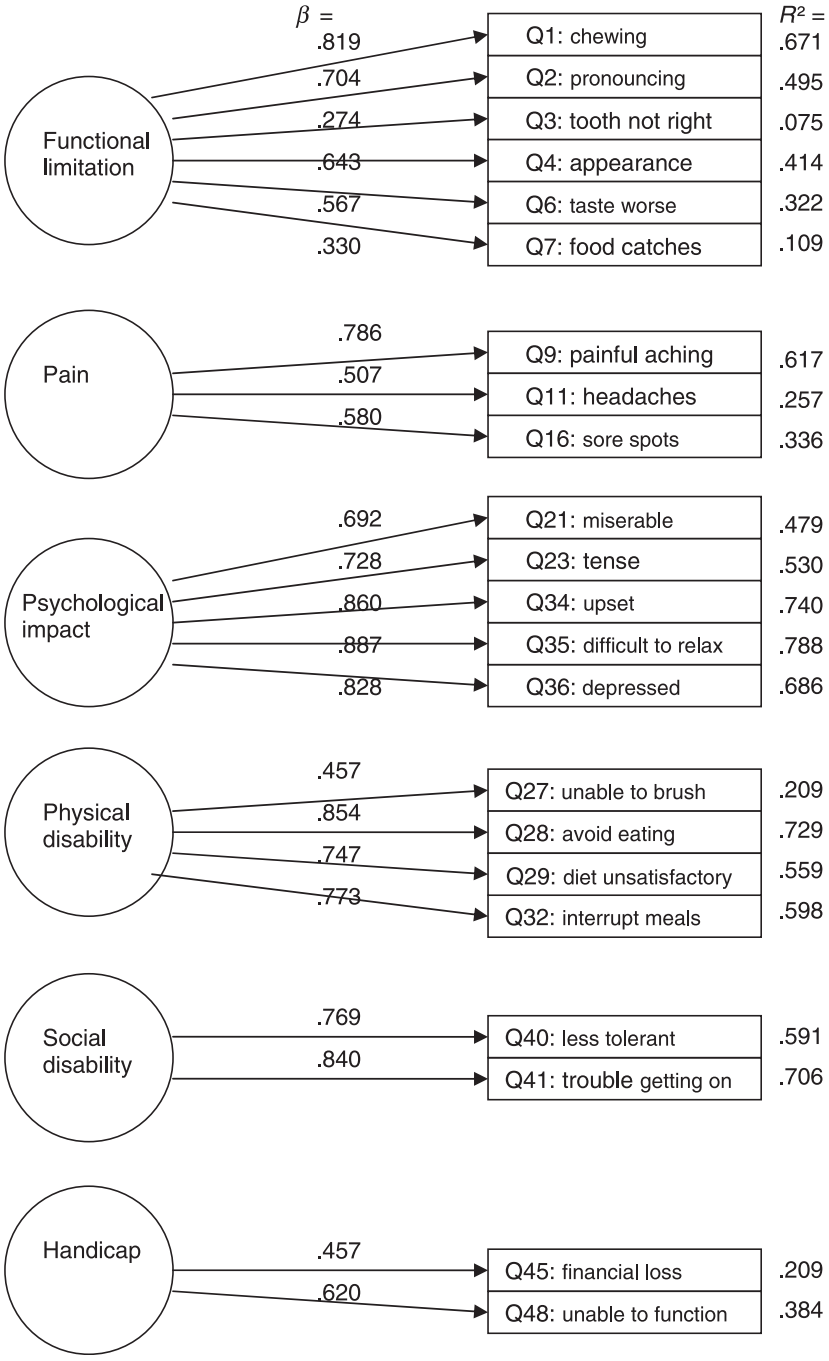

Fig. 2. Confirmatory factor analysis model of the six-factor 22 item reconceptualized OHIP scale.

relationships between the constructs as hypothesized in Locker's conceptual model of oral health (see Fig. 1). In accordance with the model, we hypothesized that functional limitation would predict disability and psychological impact (i.e. discomfort/psychological disability combined) which, in turn, would be associated with handicap. Additionally, that functional limitations would predict handicap, and pain would predict disability and psychological impact.

This model did not fit the data well (χ^2 = 833.035 (199), $P < 0.001$, χ^2/d.f. = 4.186, TLI = 0.863, CFI = 0.882, RMSEA = 0.077). Within SEM, modification indices (MIs) indicate possible changes that could be made to make the model better 'fit' the data. That is, they give an approximation of how much change to model fit (χ^2) there would be if certain parameters (e.g. covariances, regression effects) were allowed to be freely estimated. In our model, the MIs indicated that if functional limitations were allowed to predict pain, this may help the fit of the model (par change = 0.544). Given that conceptually, it is likely that restrictions in bodily functions are linked to the experience of pain, and that this pathway has been supported in previous CFA research with OHIP14 (8), the model was

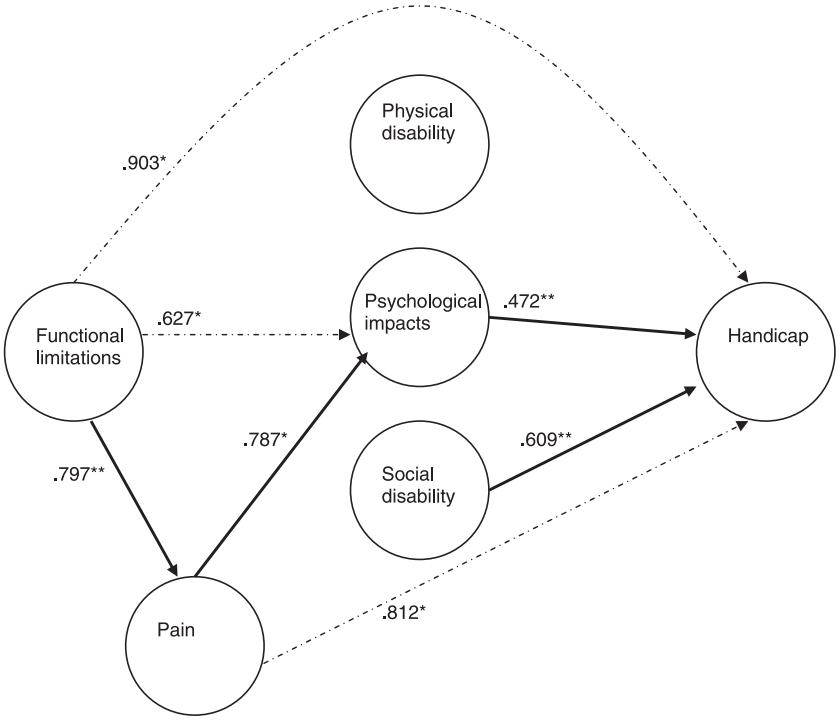

Fig. 3. Bootstrapped standardized estimates for the structural six-factor 22 item Locker model.

re-run with the additional path added. The model resulted in a significantly better fit ($\Delta\chi^2$ = 175.038 (1), $P < 0.001$) (see Table 2, Model 5). The variance accounted for within each construct was 64% (pain), 76% (psychological impact), 86% (physical disability), 31% (social disability) and 82% (handicap). As can be seen from Fig. 3, many of the direct relationships hypothesized within Locker's original model were not supported. Greater functional limitations and pain were not predictive of more physical or social disability; physical disability was not linked to handicap. In addition, other relationships hypothesized within the model were indirect rather than direct. Functional limitations were linked indirectly to psychological impacts (via pain) and to handicap (via pain and psychological impacts); pain was linked indirectly to handicap (via psychological impacts).

In summary: The between-construct validity of the re-specified 22 item OHIP was not adequately supported by the data presented here. The relationships between the six constructs were only partially as hypothesized within Locker's conceptual model.

Discussion

The present study represents the first test of the conceptual basis of the original OHIP49 since its development over a decade ago (4). Our findings suggest that the scale as currently conceived does not have adequate construct validity. There are four possible explanations for these results. First, the underlying model on which the OHIP is predicated is a scientific framework rather than empirically testable. Second, the concepts within the OHIP are broadly defined and overlapping and thus, difficult to operationalize. Third, items within the OHIP do not represent their underlying concepts. Fourth, our sample of community dwelling elders with low oral health impacts did not provide an adequate test of the model.

When originally developed, the OHIP was said to be derived from Locker's conceptual model of oral health (6). Yet, as evidenced here, many items within a construct were found to be redundant, in that they were conceptually so similar as to add little unique variance. For example, Items 48 'unable to function' and 49 'unable to work'. Using a few good indicators of a construct is generally better than including multiple items which are diffuse and unfocused (15). Furthermore, many items appeared to measure more than one construct. For example, Item 38 'been embarrassed' originally developed as a measure of 'psychological disability' also correlated highly with items from functional limitations (2 'trouble pronouncing' and 4 'appearance affected'), psychological discomfort (20 'self-conscious' and 22 'uncomfortable due to appearance), physical disability (31 'avoid smiling') and social disability (39 'avoided going out' and 40 'less

tolerant of others'). It would appear that the original labelling of items as indicators of 'functional limitation' or 'physical pain' without any further construct validation has resulted in a measure that incorporates constructs with different names but which have similar *meanings*. When constructs are not discrete they become difficult to interpret and theoretically meaningless (16). For example, the item content of the constructs 'psychological discomfort' (Items 21 and 23) and 'psychological disability' (Items 34, 35 and 36) was so similar that having two separate factors could not be supported by the data presented here.

Such findings have a number of important conceptual and practical implications. Most notably, it is questionable whether the OHIP49 should be used as a measure of Locker's conceptual model of oral health. If the OHIP were to be used as the 'best' current measure of oral health functioning, separate sub-scale scores should not be reported and, most importantly, clinicians and researchers should not distinguish between different OHIP dimensions in the analysis of treatment effects. This is because, given the lack of adequate construct validity, it is not safe to assume that the different dimensions (pain, disability, etc.) measure what they say they do.

The next step for future research is to carry out cross-validation analysis of the present results (8). In our sample of community dwelling older people, we found limited support for Locker's conceptual model as originally specified. However, given that previous research testing the model with the OHIP14 in a general UK adult population was supportive (8), further examination of the underlying model in different populations is necessary before firm conclusions can be drawn. It may be that in a nonclinical population with relatively low impact scores (such as in the present study) some items are poor indicators of their hypothesized constructs; however, if assessed in a clinical group with more severe oral health impacts, items may have better within-construct validity.

What is clear at this stage, however, is that firstly, oral health impacts are multi-dimensional as opposed to uni-dimensional and secondly, many of the relationships within the model may be indirect rather than direct. These complex interrelationships between different consequences of oral disease (impairment, disability etc) are likely to differ not only across clinical conditions but, in addition, might be modified by individual and environmental variables. The original Locker (6) model did not include such psychosocial factors, although it was noted that they were likely to play an integral role in oral health. Here, the model utilizing the six-factor 22 item reconceptualized scale explained 64% of the variance in pain, 76% in psychological impact (psychological disability and discomfort combined), 86% in physical disability, 31% in social disability and 82% in handicap. It could be that the inclusion of key contextual factors may have increased the model's explanatory power. A host of factors have been identified in the psychology literature, upon which dentistry could draw. These include coping strategies (17), social support (18), sense of coherence (19) and negative affectivity (20). There is extensive evidence that negative affectivity (predisposition to experience chronic negative emotions), for example, influences symptom perception, as well as physical health reports generally (21, 22) and more specifically OHQoL (23). Similarly, sense of coherence (a measure of the salutogenic resources available to an individual) has been shown to act as a mediator between disability and handicap (19). In addition to incorporating key contextual factors, future research on the model should engage with the revised WHO framework on which Locker's original model was predicated, namely the International Classification of Functioning, Disability and Health (24), as well as contemporary debates both within the disability literature on the ICF (25) and quality of life (26).

Despite Locker's model being published nearly two decades ago, it has received little further conceptual development. Yet, such work is vital in order to enable a more complex appreciation of oral health which will, in turn, facilitate treatment planning and the development of effective programmes to improve OHQoL, as well as wider well-being. In order to explore such processes future studies need to be of a longitudinal design. Whilst we modelled our data based on the causal ordering hypothesized within Locker's model, such ordering does not imply a causal effect (27). We focussed on cross-sectional data, and tested only unidirectional paths between variables. Utilizing longitudinal designs would allow examination of bi-directional or reciprocal relationships. It may be, for example, that an individual's perception of pain influences subjective appraisals of well-being (discomfort) or the ability to perform daily activities (disability) or *vice versa*. In addition, future studies should incorporate appropriate statistical modelling techniques which will aid the

development of theoretically-based measures and allow the testing of complex interrelationships, including both direct and indirect (mediated) effects. Only by testing such mediation models will it be possible to gain a greater understanding of the complexity of the causal processes underpinning oral health impacts. For example, in this study, functional limitations impacted on handicap but this effect was indirect via pain and psychological impacts. Translating this result into intervention strategies aimed at minimizing handicap resulting from oral disease, would suggest targeting, through physical and/or psychological intervention, individual's perceptions of pain, discomfort and psychological disability (e.g. headaches, self-consciousness, tension).

Conclusions, caveats and implications

Our findings indicate that the within- and between-construct validity of the OHIP in its current form is questionable, and that the scale needs to undergo further testing before it can be assumed that it measures the concepts (and model) as originally proposed. Nevertheless, we tested only a small number of possible structural equation models with only one sample and using secondary data analysis. There may be a number of other equally valid alternative models (28). The present model and possible alternatives now need to be cross-validated in other samples. Such cross-validation should, where possible, be based on primary data collected for the purpose of testing Locker's conceptual model rather than rely on secondary analysis with its associated limitations.

In sum, the present study extends the original paper reporting development of the OHIP, and provides a heuristic for researchers testing the validity of existing measures, as well as providing a template for development of new measures based on *a priori* theoretical models. It has highlighted some of the conceptual and methodological confusions within the OHQoL literature, and supports the need to move away from simple descriptive or correlational research in an attempt to understand the complexity of oral health from within a patient-centred perspective.

References

1. Gift HC, Atchison KA, Dayton CM. Conceptualising oral health and oral health-related quality of life. Soc Sci Med 1997;44:601–8.
2. Allen PF, McMillan AS, Locker D. An assessment of sensitivity to change of the oral health impact profile in a clinical trial. Community Dent Oral Epidemiol 2001;29:175–82.
3. Awad MA, Locker D, Korner-Bitensky N, Feine JS. Measuring the effect of intra-oral implant rehabilitation on health-related quality of life in a randomised controlled clinical trial. J Dent Res 2000;79:1659–63.
4. Slade GD, Spencer AJ. Development and evaluation of the oral health impact profile. Community Dent Health 1994;11:3–11.
5. Slade GD. Derivation and validation of a short-form oral health impact profile. Community Dent Oral Epidemiol 1997;25:284–90.
6. Locker D. Measuring oral health: a conceptual framework. Community Dent Health 1988;5:3–18.
7. Hoyle RH, Smith GT. Formulating clinical research hypotheses as structural equation models – a conceptual overview. J Consult Clinical Psychol 1994; 62:429–40.
8. Baker SR. Testing a conceptual model of oral health: A structural equation modelling approach. J Dent Res 2007;86:708–12.
9. Payne B, Ford J, Locker D. Loss to follow-up in a longitudinal oral health survey of older adults. Community Dent Oral Epidemiol 1995;23:297–302.
10. Arbuckle JL. Amos 6.0 User's Guide. PA, USA: Amos Development Corporation; 2005.
11. Kline RB. Principles and practice of structural equation modelling. 2nd edn. New York: The Guildford Press; 2005.
12. Brown TA. Confirmatory factor analysis for applied research. New York: The Guildford Press; 2006.
13. Shrout PE, Bolger N. Mediation in experimental and nonexperimental studies: New procedures and recommendations. Psychol Methods 2002;7:422–45.
14. Hu L-T, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Model 1999;6:1–55.
15. Ratner PA, Bottorff JL, Johnson JL, Hayduk LA. Using multiple indicators to test the dimensionality of concepts in the health promotion model. Res Nurs Health 1996;19:237–47.
16. Anderson JC, Gerbing DW. Structural equation modelling in practice: a review and recommended two-step approach. Psychol Bull 1988;103:411–23.
17. Baker SR. Toward an idiographic understanding of the role of social problem solving in daily event, mood, and health experiences: A prospective daily diary approach. Brit J Health Psychol 2006;11: 513–31.
18. Cohen S, Wills T. Stress, social support, and the buffering hypothesis. Psychol Bull 1985;98:310–57.
19. Schnyder U, Buchi S, Morgeli H, Sensky T, Klaghofer R. Sense of coherence – a mediator between disability and handicap? Psychother Psychosom 1999;68: 102–10.
20. Kolk AM, Hanewald GJFP, Schagen S, Gijsbers van Wilk CMT. A symptom perception approach to common physical symptoms. Soc Sci Med 2003;57:2343–54.
21. Baker SR. Dispositional optimism and health status, symptoms, and behaviours: Assessing idiothetic

relationships using a prospective daily diary approach. Psychol Health 2007;22:431–55.
22. Pennebaker JW. The psychology of physical symptoms. New York: Springer-Verlag; 1982.
23. Kressin NR, Reisine S, Spiro A III, Jones JA. Is negative affectivity associated with oral quality of life? Community Dent Oral Epidemiol 2001;29:412–23.
24. World Health Organisation. International classification of functioning disability and health. Geneva: WHO; 2001.
25. Imrie R. Demystifying disability: a review of the International Classification of functioning, disability and health. Sociol Health Illness 2004;26:287–305.
26. Phillips D. Quality of life: concept, policy and practice. London: Routledge; 2006.
27. Holland PW. Causal inference, path analysis, and recursive structural equation models. Sociol Methodol 1988;18:449–84.
28. Bollen KA. Structural equations with latent variables. New York: Wiley; 1989.
29. Byrne BM. Structural equation modelling with AMOS. London: Lawrence Erlbaum Associates; 2001.