

Cross-cultural adaptation of the Oral Health Impact Profile (OHIP) for the Malaysian adult population.

R. Saub¹, D. Locker², P. Allison³ and M. Disman⁴

¹Community Dentistry, Faculty of Dentistry, University of Malaya, Malaysia; ²Community Dentistry, Faculty of Dentistry, University of Toronto, Ontario, Canada; ³Faculty of Dentistry, McGill University, Montreal, Canada; ⁴Public Health Sciences, Faculty of Medicine, University of Toronto, Ontario, Canada

Objective: The aim of this project was to develop an oral health related-quality of life measure for the Malaysian adult population aged 18 and above by the cross-cultural adaption the Oral Health Impact Profile (OHIP). **Method:** The adaptation of the OHIP was based on the framework proposed by Herdman *et al* (1998). The OHIP was translated into the Malay language using a forward-backward translation technique. Thirty-six patients were interviewed to assess the conceptual equivalence and relevancy of each item. Based on the translation process and interview results a Malaysian version of the OHIP questionnaire was produced that contained 45 items. It was designated as the OHIP(M). This questionnaire was pre-tested on 20 patients to assess its face validity. A short 14-item version of the questionnaire was completed by 171 patients to assess the suitability of the Likert-type response format. Field-testing was conducted in order to assess the suitability of two modes of administration (mail and interview) and to establish the psychometric properties of the adapted measure. **Results:** The pre-testing revealed that the OHIP(M) has good face validity. It was found that the five-point frequency Likert scale could be used for the Malaysian population. The OHIP(M) was reliable, where the scale Cronbach's alpha was 0.95 and the ICC value for test-retest reliability was 0.79. Three out four construct validity hypotheses tested were confirmed. OHIP(M) works equally well as the English version. **Conclusion:** OHIP(M) was found to be reliable and valid regardless of the mode of administration. However, this study only provides initial evidence for the reliability and validity of the measure. Further study is recommended to collect more evidence to support these results.

Key words: Cross-cultural adaptation, OHIP, oral health-related quality of life,

Introduction

It is well accepted now that the measurement of disease alone is not sufficient to describe the oral health status of individuals or populations or when comparing the efficacy and effectiveness of interventions. The reason for this is that thinking about health and oral health has changed. Health is no longer seen as the absence of disease but rather in terms of obtaining or maintaining optimal functioning and social and psychological well-being (Locker, 1997). Thus, there are a growing number of investigators who are developing ways of measuring the impact of oral diseases on the well being of individuals and communities. As a result, several instruments have been developed to evaluate oral health related quality of life and most of these measures were developed in English speaking countries and in the English language. Thus, they cannot be used in other countries with a different language and cultural background without some form of cross-cultural adaptation.

In general, there are two ways to develop a health-related quality of life instrument for a different culture (Guillemin *et al*, 1993): to develop a totally new measure or modify a previous measure through a cross-cultural adaptation process. The first option is a time consuming process in which the bulk of the effort is devoted to the conceptualization of the measure and the selec-

tion and reduction of items. In the second option, a transposition of a measure from its original context to the target population through translation and adaptation is undertaken. This option is rather cheaper and less time-consuming as compared to the first. This method is advocated since it facilitates the use of the instrument in international and cross-cultural studies. In addition, cross-cultural adaptation not only allows us to describe or evaluate the health status of the target population, but also allows us to make comparisons between cultures or nations and the differences and similarities resulting from the various health care systems, the differing attitudes of health professionals and patients' health care behaviors (Touw-Otten and Meadows, 1996).

However, the process of adaptation is not a direct or simple task. Since culture has a significant impact on an individual's life, the measurement of health cannot be culture free. Thus, several guidelines have been proposed on how to adapt health status measures for cultures other than the one in which they were developed.

There are a number of OHRQoL measures, such as the Oral Health Impact Profile (Slade and Spencer, 1994) and the Oral Impacts on Daily Living (Adulyanon and Sheiham, 1997), that have been translated into different languages (Allison *et al*, 1999; Tsakos *et al*, 2001; John *et al*, 2002; Wong *et al*, 2002a; Wong *et al* 2002b; Tubert-Jennin *et al*, 2003) using different methods. How-

ever, most of these studies did not utilize a framework to guide the adaptation process; rather they concentrated on the translation process only. Since adaptation involves more than translation, such an approach may result in a questionnaire that addresses issues that are irrelevant to the target population or omits issues which are important to the target population.

The aim of the study described here was to develop an oral health-related quality of life measure for the Malaysian adult population (those who aged 18 and above) through the cross-cultural adaptation of the Oral Health Impact Profile (OHIP) (Slade and Spencer, 1994). The OHIP was chosen since it is a comprehensive measure that has been used in a variety of language and cultural settings. The protocol of the study and all study procedures were reviewed and approved by the Faculty of Dentistry Research Committee and the Health Sciences Committee of the Ethics Research Office at the University of Toronto.

Methods

The methodological framework

Cross-cultural adaptation is a process of producing an equivalent measure for quantifying a similar phenomenon in different cultures. Many researchers have produced guidelines by means of which cross-cultural equivalence can be achieved. In this study we used the framework proposed by Herdman *et al* (1998). We chose this model because it addresses every aspect of the questionnaire, such as conceptual basis, items and response formats and provides clear guidelines on how to achieve and assess equivalence. This model includes six types of equivalence; that is conceptual, item, semantic, operational, measurement and functional equivalence. The definition of each type of equivalence and methods by means of which each type of equivalence can be assessed is shown in Table 1.

The OHIP

OHIP was developed in Australia by Slade and Spencer (1994). It contains 49 items grouped into seven subscales; namely functional limitation (e.g. difficulty chewing), physical pain (e.g. toothache), psychological discomfort (e.g. self consciousness), physical disability (e.g. changes in diet), psychological disability (e.g. reduced ability to concentrate), social disability (e.g. avoiding social interaction) and handicap (e.g. being unable to work productively). The number of items in each domain varies. A five-point frequency Likert scale is used as the response format and treated both as ordinal and interval data. The time reference used is one year. Each item is given a weight indicative of the severity of the problem it describes. Unweighted and weighted scores can be computed for each domain and the OHIP overall. The design of the OHIP permits it to be administered both by interview and self-completed questionnaire.

The translation process

The aim of this stage was to produce a Malay version of the questionnaire, using the forward-backward translation technique to establish semantic equivalence (Behling and Low, 2000; Del Greco *et al*, 1987). Three bilingual forward translators (FT) working independently translated the original English questionnaire (OQ) into the Malay language. Prior to the translation process, the translators were briefed by the investigator (RS) on the aims of the instrument and the population on which it would be used. They were also asked to note any items, which were difficult or impossible to translate. Following the forward translation, discussions between the investigator and the three translators were undertaken to achieve one single Malay version (MQ). Three other bilingual translators then back translated this Malay version into English. The original version was not given to the back-translators to avoid bias in their back translation. The same process as in the forward translation was carried

Table 1. Definition of equivalence (Herdman *et al*, 1998)

<i>Type of equivalence</i>	<i>Definition</i>	<i>Method</i>
Conceptual	Ways in which different populations conceptualize health and quality of life (QoL) and the values they place on different domains of health and QoL.	<ul style="list-style-type: none"> • Local literature • Consultation with expert • Interview with target population
Item	Concerns the way in which domains are sampled. Item equivalence exists when items estimate the same parameters on the latent trait being measured and when they are equally relevant and acceptable in both cultures.	<ul style="list-style-type: none"> • Review literature • Expert judgment • Interview with target population
Semantic	Concerned with the transfer of meaning across languages.	<ul style="list-style-type: none"> • Forward-backward translation • Check by lay panel
Operational	Refers to the possibility of using a similar questionnaire format, instructions, mode of administration, and measurement method (response format).	<ul style="list-style-type: none"> • Pretest the questionnaire
Measurement	Ensuring that different language versions of the same instrument achieve acceptable levels in terms of their psychometric properties – reliability, responsiveness, and validity.	<ul style="list-style-type: none"> • Reliability and validity study on the target population
Functional	The extent to which an instrument does what it is supposed to do equally well in two or more cultures.	<ul style="list-style-type: none"> • Examine other equivalence

out to produce a single back-translated English version (BTQ). A committee comprised of the investigator and the six translators, reviewed the translation version in terms of: 1) the comprehensiveness of the translation. This includes the translation of the introduction and instructions, and 2) semantic equivalence. This was to ensure that items, which have been modified during the translation process, retained their original meaning. Once the committee was satisfied, the back translation was then assessed for any discrepancies with the original English language version by a panel of evaluators whose first language was English.

The evaluators comprised the developer of the original version (Dr. Slade), and two co-investigators on the project (DL and PA). Items which were considered by the evaluators as problematic, were brought to the translation committee for discussion and amendments were made. This process continued until the evaluators were satisfied with the result.

Qualitative Interviews

The aim of this stage was to investigate the conceptual and item equivalence between the source and target culture with respect to oral health, to ensure that the dimensions comprising the OHIP were appropriate and comprehensive and to identify impacts/problems not covered by the OHIP. A semi-structured in-depth interview was conducted with 36 patients attending clinics at the Faculty of Dentistry, University of Malaya. The focus of the interview was on the ways in which their oral problems impact on daily life and psychosocial well-being. After this component of the interview was completed, the Malay translation of the OHIP was shown to each patient who was asked to comment on the relevance of each item (i.e. does the problem it describes apply to them) and whether or not the statement is clear or unclear in terms of its meaning. The interviews were tape-recorded.

Each of the interviews was transcribed in full and the transcription checked against the tape for accuracy. Once the accuracy of the transcription was ensured, a coding and sorting process was undertaken using the Ethnograph software. Words or phrases that describe the functional and psychosocial impacts of oral disorders were identified and a code was inserted into the text. The codes were assigned according to the domains from the original questionnaire, as shown in Table 2. The data were then sorted according to the domains. If additional domains were identified, these would be given codes.

Table 2. Codes for Oral Health Impact Profile domains

<i>Domain</i>	<i>Code</i>
Functional limitation	FL1
Physical pain	P1
Psychological discomfort	P2
Physical disability	D1
Psychological disability	D2
Social disability	D3
Handicap	H1

Committee review

Revision of the content of the questionnaire based on a list of impacts/problems obtained from the qualitative interview and patients' comments concerning the relevance and clarity of the OHIP items was carried out. The committee that developed the translated versions of the questionnaire reviewed these revisions.

Pretest

The aim of this stage was to check the wording, clarity, as well as comprehensibility of the adapted questionnaire since it had undergone some modification (new items were added and some items were modified). This stage was also used to check whether respondents were interpreting the meaning of each item in an appropriate manner. A convenience sample of 20 respondents was selected from the patients attending clinics at the Faculty of Dentistry, University of Malaya. Each patient was given a copy of the translated Malay OHIP version to complete. After the patient had completed the questionnaire, he/she was asked to comment on the questionnaire as a whole. Then random questions were selected and the interviewer probed by asking the patient a question: "What do you think is meant by this question?" Patients were encouraged to elucidate their understanding of the items in an open-ended manner. Patients' comments were recorded verbatim. Data from the questionnaires and interviews were analyzed and any necessary changes to the wording of items, instructions or response formats were carried out.

Testing of Response options.

The aim of this stage was to test whether the response format of the questionnaire was appropriate for a Malaysian population. The response format was a five-point Likert frequency scale, with the options 'very often', 'fairly often', 'occasionally', 'hardly ever' and 'never'. A convenience sample of 171 patients attending dental clinics at the Faculty of Dentistry, University Malaya and Bangsa dental clinic was selected. To test the response options, a short 14-item version of the questionnaire was used. The reason for using a short version rather than long version is because the short version consists of the most common items reported, therefore it was expected that the respondents would be more likely to use the full range of options. Patients were asked to complete a self-administered version of the questionnaire. Data were analyzed to determine if the respondents used the full range of response options by plotting the frequency distribution for each item.

Field Testing

At this stage the full 45 item version of the OHIP(M) was used. Two modes of administration, mail versus interview, were assessed and the reliability and validity of the questionnaire were evaluated. The results of the psychometric assessment were then compared with similar assessments undertaken in Australia and Canada in order to assess the measurement equivalence of the OHIP(M).

A cross-sectional, population based sample, study design was used in this component of the study. The

participants were a sub-sample of the Malaysian National Oral Health Survey of Adults (NOHSA 2000). A total of 220 respondents were involved in this study. Two methods of administration, mail questionnaire and interviews, were employed. In order to carry out test-retest analysis, a second administration was carried out on a selected sub-sample fifteen days after the first. The same method of administration as at the first administration was employed. At this stage one additional question regarding whether the participant's oral health had changed since the first administration was added.

Two procedures were applied in the case of missing data (blank entries or "don't know" responses): 1) total exclusion - If nine or more items were missing, then the subject was excluded from the final analysis, and 2) mean item imputation - In the case where there were less than nine items with missing data, the value was imputed using the mean of that particular item (Slade, 1997).

Two methods of scoring were computed: 1) Additive (ADD score) - calculated by adding up the response codes for each item, and 2) Simple count (SC score) - calculated by summing the number of items reported as "very often" and "often" (Allen and Locker, 1997). The ADD score could range from 0 to 180 and the SC score from 0 to 45. High scores indicated poorer OHRQoL.

Appropriate statistical analyses were performed. A *p* value was set at 0.05. The analyses performed were:

- *To compare the two modes of administration (Interview vs mail)* - The comparison of the two types of mode of administration was made by assessing three parameters: response rates, completeness of data and OHIP(M) scores. A questionnaire was considered incomplete if more than 20% of responses were left blank or marked don't know (Slade and Spencer, 1994). Differences between modes of administration were tested using the χ^2 test or Mann-Whitney test.
- *Reliability* - Two types of reliability were assessed: internal consistency and test-retest reliability. Internal consistency was assessed using Cronbach's reliability coefficient α based on the responses from the first administration. Scores of 0.6 or more indicates good to excellent reliability (Locker and Slade 1993). To ensure that the questionnaire was reproducible in stable subjects, intraclass correlation co-efficients (ICC) were calculated using scores from the repeated administrations of the OHIP(M). An ICC of 0.7 was considered as an acceptable level (Aday, 1996).
- *Validity* - Data from the first administration of this part of the study was used to assess the validity of the instrument. Because there was no "gold standard", construct validity was assessed. The following hypotheses were tested: 1) Those who perceived their oral health as either very good or good would have lower OHIP(M) score; 2) Those who were not satisfied with their oral health would have higher OHIP scores; 3) Those who perceived that they needed dental treatment would have higher OHIP scores than those who did not, and 4) Edentulous subjects and dentate subjects wearing dentures would have higher OHIP scores than dentate subjects not wearing dentures. Dentate no denture means dentate people with no removable denture.

- *Comparison of the technical properties of the OHIP(M) with OHIP(Australia), and OHIP(Canada)* - The results of the reliability and validity analyses were compared with those from the OHIP (Australia) (Slade and Spencer, 1994) and OHIP (Canada) (Locker and Slade, 1993) to establish measurement equivalence. The Cronbach's values and intraclass correlation coefficients were compared. Two analyses were performed in order to compare the ability of the measures to distinguish between groups: the differences in the mean ranks (DMR) between categories of the independent variables, obtained from the Mann-Whitney test and odds ratios based on median splits (Allen and Locker, 1997).

Results

Translation

Thirty-seven out of 49 OHIP questions were rated as easy to translate and none of the items were considered impossible to translate. Twelve questions were considered as difficult to translate, predominantly those assessing socio-emotional issues. In some cases, two different emotional states in the English language were translated into the same word in the Malay language, for example, "miserable" and "upset". A similar situation was also observed by Hunt *et al* (1986). Hunt stated that "the socio-emotional items are likely to cause the most trouble and items what are perhaps more universal experiences, like pain, sleep and impaired physical mobility, the least." Nevertheless, most of the items in the original questionnaire were considered by the translators to be easy to translate.

Qualitative Interviews

Forty-nine significant impacts were identified. Of these, three were considered to be new impacts, which had frequently emerged in the interviews. They were: "unable to enjoy favorite foods", "loss of appetite" and "lacking in self-confidence". Although there were impacts reported that were not listed in the original questionnaire, they belonged to one of the original subscales. No new subscale emerged, nor was an existing subscale excluded. Thus, based on this finding it could be concluded that Malaysians and Australians conceptualize oral health similarly in terms of how oral diseases or oral disorders affect daily life.

In the second part of the interview, the patients were shown the translated questionnaire and asked to comment on the relevancy and their understanding of each statement. Most of the items were understood and relevant. Two pairs of questions in the translated version were viewed by most of the patients interviewed as being the same. They were "self-conscious" and "embarrassed", and "miserable" and "upset".

The Malaysian OHIP version.

Based on the translation process and qualitative interview results, seven items from the original OHIP were excluded either because they were ambiguous, yielded similar meanings to other items, or were not relevant. Three items were restructured and/or reworded and three

items were added. This adapted questionnaire is referred to hereafter as the Malaysian Oral Health Impact Profile [OHIP(M)], which contains 45 questions.

Pretest

All respondents agreed that the format and instructions on how to complete the questionnaire were easy to follow. Almost all respondents completed the entire questionnaire. This indicates that the instructions and the questions were easily understood. On average, the time taken to complete the OHIP(M) was 20 minutes. Most of the respondents had no difficulty understanding the questions except one person who had difficulty understanding the word “*rahang*” (jaw). Since only one person did not understand the word “*rahang*”, it was retained. No major changes were made. Thus, the questionnaire was not re-tested.

Test of response options

A short 14-items version of the OHIP(M) was used to test the response format. Figure 1 shows an example of the distribution of the response option. This indicates that the response format of a five-point frequency Likert scale can be used for a Malaysian adult population.

Field testing

Comparison of mode of administration – mail vs interview

Table 3 shows the response rate, percentage of incomplete data and the OHIP(M) scores – ADD score and SC score – by mode of administration. The interview response rate was significantly higher than the response rate for the mail questionnaire. The percentage with incomplete data for the OHIP(M) mail questionnaire was four percent and for the interview was zero percent. It was observed that the mean scores – both ADD score and SC score – were lower for interviews than for mail. However, the differences were not significant.

Reliability

Internal consistency was assessed by Cronbach’s alpha. The overall Cronbach’s alpha for the seven subscales of the OHIP(M) ranged from 0.72 to 0.87 (Table 4). A higher value was observed for the mail questionnaire for all the subscales, which ranged from 0.75 to 0.88 than the interview, which ranged from 0.66 to 0.84. However, regardless of mode of administration all values were more than 0.60, indicating good to excellent reliability. A total of 71 respondents completed the OHIP(M) for the second time. Forty-two respondents completed the mail questionnaire and 29 respondents were re-interviewed. The ICCs ranged from 0.67 to 0.80 for the seven subscales, indicating good to excellent test-retest reliability (Table 4).

Validity

It was observed that the mean score of the OHIP(M) increased as the respondents’ perceived oral health status changed from good to poor (Table 5). This supported the construct validity of the measure whereby the greater the impact on the quality of life, the poorer the perceived oral health status.

It was observed that those respondents who reported that they did not need dental treatment and were satisfied with their oral health had a lower score than those who perceived they required treatment and were not satisfied with their oral health. These differences were statistically significant.

Table 6 shows the mean score by dental status of the OHIP(M) for the seven subscales. The edentulous respondents had a higher score for both ADD and SC on the OHIP(M) scales than the dentate with dentures and the dentate without dentures. On the other hand, the dentate without dentures had the lowest score for both ADD and SC scores. However, these differences were not statistically significant. When examined for each of the subscales, there was a statistically significant difference in scores, both for ADD and SC, between the three groups on the functional limitation and physical disability subscales.

Comparison of the technical properties of the OHIP(M) with the Australia OHIP(A) and Canadian data OHIP(C)

Generally, the internal consistency of the OHIP(M) was better than that of the OHIP(A), especially for the handicap domain. However, the OHIP(A) had better internal consistency reliability than the OHIP(M) for the pain domain. In contrast, the OHIP(C) had better internal consistency for most of the subscales than the OHIP(M). In terms of stability, two subscales of the OHIP(M) had an ICC slightly lower than 0.7 compared to the three subscales of the OHIP(A) that were lower than 0.5. This suggests that the OHIP(M) was a more stable instrument than the OHIP(A) (Table 7).

As shown in Table 8, differences in mean ranks (DMR) obtained from Mann-Whitney tests indicated that both measurements discriminate between groups; however, the OHIP(C) performed better in this regard. When the odds ratio based on median splits was calculated, the OHIP(M) seemed to perform marginally better than the OHIP(C). Since the database of the OHIP(A) could not be accessed, the same analysis could not be performed on the OHIP(A).

Discussion

The aim of this study was to produce an oral health-related quality of life measure to be used in Malaysian adult population. Since there are number of such measures available, we decided to adapt the OHIP, the most commonly used of those measures. In this study a comprehensive cross-cultural adaptation was carried out based on the framework proposed by Herdman *et al* (1998). Herdman *et al*’s framework provides a set of clear guidelines for adapting a measure and assessing cross-cultural equivalency. While translation begins the process of adaptation additional procedures are necessary to ensure that a questionnaire is suitable for and works equally well in a different culture. This study has archived all the equivalence assessed as summarized in Table 9

The OHIP was translated into the Malay language, since it is the national language of Malaysia. In order to make the content culturally sensitive to the target population some items were deleted, rephrased or added based on the results of one or other of the sub-studies

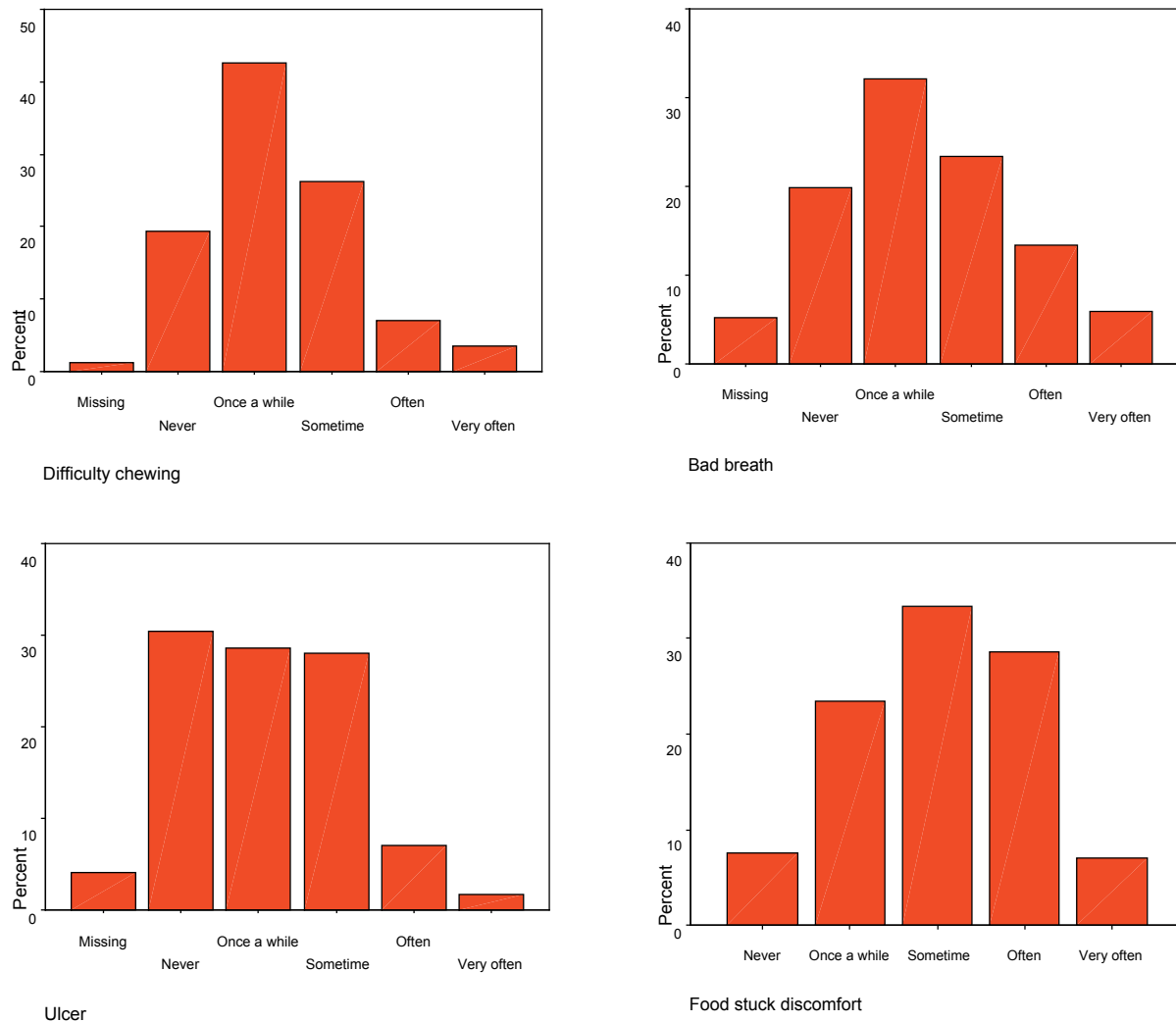


Figure 1. Distribution of response options for some of the questions.

Table 3. Response rate, percentage of incomplete data and the OHIP(M) scores – ADD score and SC score by mode of administration

	Response Rate <i>n (%)</i>	Incomplete Data <i>n (%)</i>	OHIP(M) Score	
			ADD-score Mean (SD)	SC-score Mean (SD)
Mail	149 (48.1)*	6 (4.0)	28.48 (24.78)	2.71 (4.95)
Interview	71 (56.8)*	0 (0.0)	25.27 (18.47)	1.89 (3.13)
TOTAL	220 (50.6)	6 (2.7)	27.42 (22.89)	2.44 (4.44)

* χ^2 test; $P < 0.05$

Table 4. Cronbach's alpha and ICC values.

Subscales (No. of Items)	Cronbach's alpha			ICC		
	Mail (<i>n</i> =143)	Interview (<i>n</i> =71)	Total (<i>n</i> =214)	Mail (<i>n</i> =42)	Interview (<i>n</i> =29)	Total (<i>n</i> =71)
Functional Limitation (6)	0.77	0.73	0.76	0.79	0.79	0.79
Physical Pain (7)	0.75	0.66	0.72	0.78	0.81	0.79
Psychological Discomfort (6)	0.85	0.80	0.84	0.64	0.74	0.67
Physical Disability (9)	0.88	0.84	0.87	0.71	0.73	0.71
Psychological Disability (6)	0.88	0.73	0.86	0.75	0.59	0.72
Social Disability (5)	0.83	0.73	0.81	0.84	0.63	0.77
Handicap (6)	0.82	0.66	0.78	0.75	0.84	0.80
Scale (45)	0.96	0.93	0.95	0.76	0.84	0.79

Table 5. Mean ADD-Scores by self rated oral health

	<i>Very good/Good</i> (<i>n</i> =105)	<i>Fair</i> (<i>n</i> =102)	<i>*p</i>
Functional Limitation	2.98 (3.59)	5.97 (3.97)	0.000
Physical Pain	4.80 (3.66)	7.37 (3.84)	0.000
Psychological Discomfort	3.56 (3.67)	7.22(4.68)	0.000
Physical Disability	3.27 (4.82)	6.65 (5.75)	0.000
Psychological Disability	1.95 (2.97)	4.46 (4.18)	0.000
Social Disability	0.85 (1.89)	1.75 (2.53)	0.000
Handicap	1.45 (2.15)	3.46 (3.37)	0.000
Scale	18.87 (18.48)	27.74 (22.96)	0.000

* Mann-Whitney test.

Table 6. Mean score of the OHIP(M) by dental status.

<i>Subscale</i>	<i>Mean ADD Score (SD)</i>				<i>Mean SC score (SD)</i>			
	<i>Dentate</i> <i>no denture</i> (<i>n</i> =150)	<i>Dentate</i> <i>with denture</i> (<i>n</i> =49)	<i>Edentulous</i> (<i>n</i> =15)	<i>*p-value</i>	<i>Dentate</i> <i>no denture</i> (<i>n</i> =150)	<i>Dentate</i> <i>with denture</i> (<i>n</i> =49)	<i>Edentulous</i> (<i>n</i> =15)	<i>*p-value</i>
Functional Limitation	3.83 (3.46)	5.18 (4.53)	7.73 (5.43)	0.007	0.37 (0.78)	0.57 (1.09)	1.27 (1.49)	0.020
Physical Pain	5.97 (4.03)	6.59 (3.61)	5.57 (4.06)	0.430	0.37 (0.88)	0.43 (0.91)	0.60 (1.12)	0.675
Psychological Discomfort	5.15 (4.23)	5.67 (5.51)	6.67 (5.01)	0.550	0.67 (1.11)	0.79 (1.47)	0.67 (1.11)	0.987
Physical Disability	3.91 (4.53)	6.67 (6.57)	9.67 (6.97)	0.000	0.31 (0.91)	0.69 (1.54)	1.60 (2.19)	0.003
Psychological Disability	3.25 (3.74)	2.75 (3.63)	4.07 (4.89)	0.594	0.17 (0.58)	0.16 (0.74)	0.53 (1.06)	0.137
Social Disability	1.27 (2.19)	1.08 (1.91)	1.93 (3.49)	0.920		0.06 (0.33)	0.20 (0.56)	0.056
Handicap	2.29 (2.75)	2.88 (3.18)	3.13 (4.21)	0.617	0.08 (0.41)	0.16 (0.55)	0.27 (0.59)	0.85
Scale	25.67 (20.59)	30.84 (25.36)	38.87 (30.26)	0.171	2.03 (3.65)	2.82 (5.34)	5.13 (6.81)	0.169

*Kruskal-Wallis test

undertaken. However, most of the OHIP items were retained because of their universality. A total of 45 questions, compared to 49 for the original English language OHIP, were finally included in the OHIP(M). The results of this and other studies that have translated or adapted the OHIP and similar questionnaires (Allison *et al*, 1999; Tsako *et al*, 2001; John *et al*, 2002; Wong *et al*, 2002a; Wong *et al* 2002b; Tubert-Jennin *et al*, 2003) suggests that there are only minor variations across cultures in the perceived impact of oral disorders.

One important aspect of a standardized questionnaire is the response format. There are several types of response format available, such as Visual Analog Scale, Likert Scale, etc. Some of them are too complicated to be used in some population groups. Thus, it is important to assess the suitability of the response format for the target population. A non-appropriate response format will result in either a loss of sensitivity or will create a "noise" on the instrument, causing a response-bias.

In this study, the suitability of the five-point frequency Likert scale response format was assessed. Because Malaysian people are very moderate when providing responses, it is necessary to test whether or not they use the entire spectrum of options. However, in this study it was observed that the response format could be used in the Malaysian adult population. Perhaps, the questions in the questionnaire were not threatening: there is evidence that the way that people respond to a question will depend on how threatening they perceive the question to be (Aday, 1996).

This study suggests that the OHIP(M) had a good internal consistency for all subscales and it was slightly better than the OHIP(A). However, it must be noted that these differences could be due to: 1) differences in the population studied (adult population aged 18+ for the present study and the elderly in Australian study), and 2) could be due to the time between the two administrations of the questionnaire (two weeks for the present study and

Table 7. Comparison of the reliability of the OHIP(M) with the OHIP(A) and the OHIP(C)

Reliability	OHIP(M)		OHIP(A) ¹	OHIP(C) ²
	18-34 (n=65)	35+ (n=149)	60+ (n=122)	50+ (n=540)
Internal Consistency (Cronbach's α):				
Functional limitation	0.68	0.77	0.70	0.84
Physical pain	0.73	0.72	0.76	0.81
Psychological discomfort	0.82	0.84	0.77	0.88
Physical disability	0.80	0.88	0.82	0.88
Psychological disability	0.83	0.87	0.83	0.90
Social disability	0.76	0.82	0.73	0.88
Handicap	0.56	0.82	0.37	0.80
Test-retest:				
Functional limitation	0.70	0.77	0.77	Not available
Physical pain	0.73	0.83	0.42	
Psychological discomfort	0.74	0.64	0.76	
Physical disability	0.83	0.66	0.72	
Psychological disability	0.73	0.71	0.74	
Social disability	0.85	0.73	0.08	
Handicap	0.77	0.78	0.48	

OHIP(M)=Malaysia, OHIP(A)=Australia, OHIP(C)=Canada.

Source: ¹Slade GD, Spencer AJ. Development and evaluation of the Oral Health Impact Profile. Community Dental Health 1994; 11:3-11.

²Secondary data from Ontario Study of the Oral Health of Older Adults database.

Table 8. Comparison of the validity of the OHIP(M) with the OHIP(C)

Construct validity	Standardized ADD score Median		Standardized SC score Median	
	M	C	M	C
Perceived oral health status				
Very good/good	7.22	7.14	0.00	0.00
Fair/poor	18.89	16.84	4.44	2.04
P-value	0.000	<0.000	0.000	<0.0001
DMR	55	142	47	85
Odds ratio (good/poor)	5.1	5.2	7.7	2.5
Perceived dental treatment need				
Need treatment	14.44	13.77	2.22	2.04
Do not treatment	5.56	7.14	0.00	0.00
P-value	0.000	0.000	0.001	0.000
DMR	52	100	27	64
Odds ratio (no/Yes)	3.5	3.2	3.2	2.1
Satisfied with oral health				
Yes	8.33		0.00	
No	20.00		4.44	
P-value	0.000	NA	0.000	NA
DMR	55		44	
Odds ratio (yes/no)	4.9		6.6	

M=Malaysia, C= Canada

P-values:- Mann-Whitney tests

DMR – difference in mean ranks between categories of grouping variables

Source: "Secondary data from Ontario Study of the Oral Health of Older Adults database.

Table 9. Summary of equivalence establishment.

<i>Semantic</i>	The semantic equivalence was ensured through a standard process of forward-backward translation and expert panel assessment.
<i>Conceptual</i>	The qualitative interviews with the target population revealed that the way Malaysian and Australian people conceptualized oral health and quality of life was similar in terms of how oral conditions affect people's lives. The domains were the same even though there were few other impacts identified.
<i>Item</i>	Item equivalence was established by recognizing items which were not relevant to the target population, removing some items which were causing some confusion, and adding new items that were considered important for the target population.
<i>Operational</i>	Operational equivalence in terms of format of the questionnaire, response options and method of administration were established. The format used has been shown to be effective since most respondents either in the pretest or the field study returned a completed questionnaire. The five-point response format can be used for the Malaysian population. Since most people in Malaysia are literate, a mail questionnaire can be used; however, the response rate could be compromised as with any other mail study.
<i>Measurement</i>	Reliability and validity were reassessed and it was found that the L-OHIP(M) was reliable and valid. It was then compared with the English version (the OHIP(A) and the OHIP(C)), and it was revealed that the L-OHIP(M) performed equally well as the OHIP(A) and the OHIP(C).
<i>Functional</i>	According to the model of equivalence (Herdman <i>et al</i> , 1998), functional equivalence is achieved when all other types of equivalence in the model have been achieved. Therefore, it can be claimed that functional equivalence has been achieved in this study since all aspects of equivalence were established.

three months for the Australian study). However, in most instances, two weeks is a sufficient time frame, especially for a long questionnaire to test for stability.

The question of how this questionnaire was going to be administered was addressed in this study. Two methods of administration were assessed: a face-to-face interview and a mail questionnaire. To assess this, the sample was divided into two groups: mail and interviews. However, it was not randomly allocated due to administrative reasons as mentioned earlier. As such, the comparison between these two modes of administration may be biased. Thus, interpretation of the results was made within this limitation.

In most surveys, mail is preferred over interviews and other methods mainly because it is cheaper. However, mail administration is not appropriate when there is a low literacy rate. In the case of the Malaysian population, statistics have shown that the literacy rate is high. Despite the high literacy rate in Malaysia, the response rate for mail questionnaire in this study was low (slightly lower than fifty percent) even when respondents were given two reminders. Perhaps the response rate could have been improved by increasing the number of reminders. However, Locker and Miller (1994) concluded that the four-wave design tends to be inefficient, with relatively few respondents responding to the fourth mailing. Due to financial constraints this study did not offer any incentives for participation, which is effective way in increasing the response rate (Edwards *et al*, 2003). Another reason could be due to using only a Malay language questionnaire. This could have led to a lower response rate among the non-Malays. Thus, perhaps a multilingual questionnaire is needed to increase the response rate among the non-Malay population, especially if a self-completed questionnaire is administered.

On the other hand, the interview yielded better response rate, as expected. However, a face-to-face interview is time-consuming and costly. In this study,

the process of finding the homes of the respondents was very time consuming, since the map was not very reliable. Furthermore, the interview could only be done on weekends because most of the respondents work on weekdays. In addition, the interview needed to be done at a very specific time in order to obtain a better response. For example, most people have lunch from 12:00 – 2:00 pm and do not like to be disturbed, and after 2:00 pm most families take an afternoon nap. Thus, perhaps the most appropriate time for interviews is from 10:00 am to 12:00 pm and after 5:00pm.

About half of the non-respondents refused to be interviewed. The main reason was because they had been involved in many studies prior to the present study. Slightly more than half of the non-respondents could not be interviewed because they were not available. Most of them did not respond to the appointment card left at their house. This could have indicated that they were not interested in participating in the study.

Incomplete data is another source of non-response bias (Streiner and Norman, 1995). In this study, the questionnaire was considered to be unusable when more than twenty percent of the items were left blank or had “don't know” responses. As anticipated, the mail questionnaire had a higher percentage of incomplete data than the interview questionnaire both for the L-OHIP(M) and the S-OHIP(M). It was also observed that the length of the questionnaire influenced the completeness of data.

Another parameter used to compare the two modes of administration was the OHIP(M) score. It was observed that the mail questionnaire had a slightly higher score than the interview; however, the differences were not significant. A crossover study conducted by Slade *et al* (1992) examined the effects of data collection methods for a self-reported instrument and concluded that “mail questionnaire responses to oral health impact questions are less prone to subject acquiescence or response bias.”

Therefore, the users of the OHIP(M) have to decide which method they want to employ. On one hand, the mail questionnaire compromised the response rate. On the other hand, it reduced the response-bias as well as the cost and time, compared to interview. It is also noteworthy that it is much easier to deal with a non-response bias than a response bias. Thus, potential users have to draw a line between response rate, cost and response bias. It is also recommended that only one method of administration should be used.

Reliability testing alone is not sufficient to establish the usefulness of a measure (Streiner and Norman, 1995). This is because reliability testing only tells us about the reproducibility of the measure but does not ensure that it measures what it was intended to measure - in other words, the scale's validity. Since OHIP(M) scores obtained from mail and interviews were not significantly different, the data was combined for validity analysis. This was done to increase the number of respondents in each category of the independent variables. Three out of the four construct validity hypotheses tested were confirmed. These findings provide evidence to support the validity of the OHIP(M).

This study demonstrated that the Malaysian version of the OHIP works equally well as the English language version. However, this study only provides initial evidence for reliability and validity of this measure. Further study is recommended to collect more evidence to support the psychometric properties of the measure when used with Malay-speaking populations.

Acknowledgement

This study was supported by Vot F: Grant no: F0369/2001c, University of Malaya, and partial fellowship award, University of Toronto.

References

- Aday, L.A. (1996). *Designing and conducting health surveys: A comprehensive guide*. Second edition. San Francisco: Jossey-Bass Publishers.
- Adulyanon, S. and Sheiham, A. (1997). Oral Impacts on Daily Performances. In: *Measuring oral health and quality of life*, ed. Slade, G.D. pp150-160. Chapel Hill, North Carolina: Department of Dental Ecology, School of Dentistry, University of North Carolina.
- Allen, P.F. and Locker, D. (1997). Do item weights matter? An assessment using the oral health impact profile. *Community Dental Health* **14**, 133-138.
- Allison, P. Locker, D. Jokovic, A. Slade, G. (1999): A cross-cultural study of oral health values. *Journal of Dental Research* **78**, 643-649.
- Behling, O. Law, K.S. (2000). *Translating questionnaires and other research instruments: problems and solution series, quantitative application in the social sciences*. Thousand Oaks: Sage Publications Inc.
- Del Greco, L. Walop, W. Eastridge, L. (1987): Questionnaire development 3: Translation. *Canadian Medical Association Journal* **136**, 817-818.
- Edwards, P. Roberts, I. Clarke, M. DiGuseppi, C. Pratap, S. Wentz, R. Kwan, I. (2003). Methods to influence response to postal questionnaires (Cochrane Methodology Review), issue 4. In: *The Cochrane Library*. Chichester, UK: John Wiley & Sons, Ltd.
- Guillemin, F. Bombardier, C. Beaton, D. (1993): Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology* **46**, 1417-1432.
- Herdman, M. Fox-Rushby, J. Badia, X. (1998): A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Quality of Life Research* **7**, 323-335.
- Hunt, S.M. McEwen, J. McKenna, S.P. (1986). *Measuring health status*. London: Croom Helm Ltd.
- John, M.T. Patrick, D.L. Slade, G.D. (2002). The German version of the Oral Health Impact Profile-translation and psychometric properties. *European Journal of Oral Sciences* **110**, 425-433.
- Locker, D. (1997). Concepts of oral health, disease and the quality of life. In: *Measuring oral health and quality of life*, ed Slade G.D pp 11-23. Chapel Hill, North Carolina: Department of Dental Ecology, School of Dentistry, University of North Carolina.
- Locker, D. and Miller, Y. (1994): Evaluation of subjective oral health status indicators. *Journal of Public Health Dentistry* **54**, 167-176.
- Locker, D. and Slade, G. (1993). Oral Health and the Quality of Life Among older adults: The Oral Health Impact Profile. *Journal of Canadian Dental Association* **59**, 830-838. (cont. ref.on page 844).
- Oral Health Division, Ministry of Health Malaysia (2004). National oral health survey of adults 2000 (NOHSA 2000). Kuala Lumpur: Oral Health Division, Ministry of Health Malaysia.
- Slade, G.D. (1997). The Oral Health Impact Profile. In: *Measuring oral health and quality of life*, ed Slade, G.D. pp 93-104. Chapel Hill, North Carolina: Department of Dental Ecology, School of Dentistry, University of North Carolina.
- Slade, G.D. and Spencer, A.J. (1994): Development and evaluation of the Oral Health Impact Profile. *Community Dental Health* **11**, 3-11.
- Slade, G.D. Spencer, A.J. Keily, P. (1992): Effects of data collection methods on self-reported oral health impact. *Journal of Dental Research* **71**, 978 Abstract 15.
- Streiner, D.L. Norman, G.R. (1995). *Health measurement scales: A practical guide to their development and use*. Second edition. Oxford: Oxford University Press.
- Touw-Otten, F. and Meadows, K. (1996). Cross-cultural issues in outcome measurement. In: *Health outcome measures in primary and out-patient care*, ed Hutchinson, A. The Netherlands: Harwood Academic Publishers.
- Tsakos, G. Marcenes, W. Sheiham, A. (2001): Evaluation of a modified version of the index of Oral Impacts on Daily Performances (OIDP) in elderly populations in two European countries. *Gerodontology* **18**, 121-130.
- Tubert-Jeannin, S. Riordan, P.J. Morel-Papernot, A. Porcheray, S. Saby-Collet, S. (2003): Validation of an oral health quality of life index (GOHAI) in France. *Community Dentistry and Oral Epidemiology* **31**, 275-284.
- Wong, M.C. Lo, E.C. McMillan, A.S. (2002a): Validation of a Chinese version of the Oral Health Impact Profile (OHIP). *Community Dentistry and Oral Epidemiology* **30**, 423-430.
- Wong, M.C. Liu, J.K.S. Lo E.C. (2002b): Translation and validation of the Chinese version of GOHAI. *Journal of Public Health Dentistry* **62**, 78-83.