Cross-cultural equivalence in translations of the oral health impact profile

Unsolicited Systematic Review

Michael I. MacEntee and Mario Brondani
Faculty of Dentistry, University of British Columbia, Vancouver, BC, Canada

Abstract – The Oral Health Impact Profile (OHIP) has been translated for comparisons across cultural boundaries. This report on a systematic search of literature published between 1994 and 2014 aims to identify an acceptable method of translating psychometric instruments for cross-cultural equivalence, and how they were used to translate the OHIP. An electronic search used the keywords ‘cultural adaptation’, ‘validation’, ‘Oral Health Impact Profile’ and ‘OHIP’ in MEDLINE and EMBASE databases supplemented by reference links and grey literature. It included papers on methods of cross-cultural translation and translations of the OHIP for dentulous adults and adolescents, and excluded papers without translational details or limited to specific disorders. The search identified eight steps to cross-cultural equivalence, and 36 (plus three supplemental) translations of the OHIP. The steps involve assessment of (i) forward/backward translation by committee, (ii) constructs, (iii) item interpretations, (iv) interval scales, (v) convergent validity, (vi) discriminant validity, (vii) responsiveness to clinical change and (viii) pilot tests. Most (>60%) of the translations involved forward/backward translation by committee, item interpretations, interval scales, convergence, discrimination and pilot tests, but fewer assessed the underlying theory (47%) or responsiveness to clinical change (28%). An acceptable method for translating quality of life-related psychometric instruments for cross-cultural equivalence has eight procedural steps, and most of the 36 OHIP translations involved at least five of the steps. Only translations to Saudi Arabian Arabic, Chinese Mandarin, German and Japanese used all eight steps to claim cultural equivalence with the original OHIP.

Key words: cross-cultural equivalence; disability; oral health; psychometric methods; social comparison; the oral health impact profile

Michael I. MacEntee, Faculty of Dentistry, University of British Columbia, 2199 Wesbrook Mall, V6T 1Z3 Vancouver, BC, Canada
e-mail: macentee@dentistry.ubc.ca

Submitted 15 April 2015; accepted 16 September 2015

Psychometric instruments aim to measure, among other attributes, human constructs or concepts of personal, social and cultural values based on concepts of shared characteristics (1). The constructs and measurements are all influenced by the beliefs of the instrument developers or respondents to the instrument’s items. Bullinger et al. (2). explain health-related quality of life (HRQoL) as an external focus on how health influences attainment of a fulfilling life in a context influenced strongly by cultural beliefs, behaviours and bias that contaminate the conceptual and colloquial translation of psychological and social expectations usually associated with the psychometrics of HRQoL. Guillemin et al. (3). complained more specifically about the lack of essential information for judging the cultural equivalence of translated instruments. Similarly, Herdman et al. (4). observed that definitions of equivalence in reports of translations,
especially ‘construct equivalence’, were frequently vague or incorrect.

The Oral Health Impact Profile (OHIP) uses 46 items identified by dental patients in Adelaide, Australia, to explain the effects of their oral impairments, limitations and disability, plus three items from the developers on how oral handicaps disturb patients. The 49 items together reflect seven themes (functional limitation, physical pain, psychological discomfort, physical disability, psychological disability, social disability and handicap) covered by the long form of the OHIP, and the responses of participants measured on a 5-point Likert scale. Others (5–7) produced shorter versions of the OHIP with 14 and 19 questions. The developers of the OHIP selected the International Classification of Impairment, Disability and Handicap (ICIDH) from an interpretation by Locker (8) of sick role theory as the conceptual foundation or construct of the OHIP to measure the personal and social consequences of oral disease, disorder and illness (9). The World Health Organization replaced the ICIDH with the International Classification of Function, Disability and Health (ICF) to link health conditions dynamically to the cultural and environmental context in which people cope with disability (10–12). It is possible therefore that the construct underlying the OHIP and the items emerging might not represent fully the potential responses that people give to oral disorder and disability in other cultures (13, 14).

Compounding the challenge to the construct validity supporting the OHIP, Allison et al. (15) raised concerns about aggregating scores from the OHIP measurements across cultural boundaries. They found that scores from English-speaking Canadians and Australians and French-speaking Canadians were reasonably consistent, but that the items might not encompass the health dimensions or even the concept of health current in other cultures (16, 17). Interpretation of questions differs between cultures and influences responses to questions about disabilities and events associated with disabilities (16, 17). Moreover, culture influences Likert scales because of the ‘reference group effect’ as people compare themselves with others in the context of different needs (18). Overall, there is general agreement that cultural characteristics influence how HRQoL is interpreted and measured but no consensus or widely supported best practice on how to translate HRQoL instruments for cross-cultural equivalence.

This study reports on a systematic search of literature to identify methods of translating psychometric instruments for cross-cultural equivalence, and how they have been used to translate the OHIP.

Methods

The systematic search of English literature published between 1 January 1994 and 23 December 2014 followed an iterative process whereby the translational methods associated with the first aim influenced the concurrent analyses of the OHIP translations (19). Essentially, before assessing translations of the OHIP, we searched for translational methods aimed at equivalence across cultures. We included papers on methods of cross-cultural translations and on translations of the OHIP for adults or adolescents with natural teeth, and excluded others that lacked details of the translational process or that were limited to specific disorders such as tooth loss or joint dysfunction.

Keywords ‘OHIP’, ‘Oral Health Impact Profile’, ‘cultural adaptation’ and ‘validation’ were used in the Medline and EMBASE databases (Table 1). Google Scholar was the primary source for reference links and grey literature, and reports on cross-cultural equivalence were identified from statements like ‘equivalent to… other countries‘ (20), ‘suitable for… multi-cultural studies‘ (21) or ‘to compare… with existing translations‘ (22). Two independent investigators read the abstracts and full texts of the papers, and reconciled differences of interpretation by mutual agreement. We followed the process of searching, sorting, selecting, acquiring, reading, identifying, refining and again searching for reference links and grey literature as recommended by Boell & Cecez-Kecmanovic (19) and the PRISMA Statement (23) until the search was saturated.

Results

The search

The electronic search produced 239 abstracts from which we rejected 153 unrelated directly to the OHIP for adults or adolescents and 50 that did not describe the translational process (Fig. 1). Full textual analyses of the remaining 36 papers identified and excluded another four without details of the translations. The 32 acceptable texts provided another 17 links to references and grey literature.
The final yield of 49 papers provided nine on methods for cross-cultural translations and 39 on translations of the OHIP including three with details on earlier translations. Altogether, there were 36 specific translations of the OHIP to 33 languages or dialects.

**Translational method**
Brislin (24) described a 4-step translational method involving forward and backward translations by multiple translators, a committee of bilingual raters to consider the equivalent meaning of each item, a pretest in the target population and finally a pilot
test among bilingual participants from the targeted culture. Others elaborated on this process by assessing in the target culture the theoretical construct or validity of the underlying concept (2–4, 25), the colloquial (2), vernacular (26) or functional interpretations of the items (25) [also called technical or concurrent validity (4) and idiomatic or experiential equivalence (3, 4, 25)], the interval scale or metric equivalence (24, 26), convergent and discriminant validity (4, 26) [also called operational equivalence (2)] and the response to clinical change (2–4, 26). The Medical Outcomes Trust in the USA emphasized the need for both conceptual and linguistic equivalence (27), and others recommended de novo methods involving personal interviews (3, 25) and focus groups (26) with laypersons to explore oral health-related experiences in the target culture.

In summary, there are eight steps relevant to cross-cultural equivalence: (i) forward–backward translations by a committee of ≥2 translators, (ii) underlying construct, (iii) item interpretations, (iv) interval scale, (v) convergent validity, (vi) discriminant validity, (vii) responsiveness to clinical change and (viii) pilot tests. Convergent or concurrent validity relates to a convergence of scores with similar items, such as association between scores and treatment needs, whereas discriminant or group validity relates to reliable discrimination between dissimilar items and the same items in different circumstances, such as association between scores with or without natural teeth (28).

The OHIP translations
Ten of the 33 languages or dialects have both long and short forms of the OHIP, and about two-thirds of the 36 translations aimed at cross-cultural equivalence (Table 2). Three-quarters of the translations were guided by a referenced source. Over two-thirds (69%) addressed the interpretation of items and interval scale and tested convergent and discriminant validity. Over half (56%) used committees or expert panels for forward–backward translations followed by pilot tests, but less than one-third tested the responsiveness to clinical change in the target population. About half of the translations assessed the underlying construct; however, only the Saudi Arabian Arabic (29), German (30, 31), Japanese (32) and Malaysian (33) versions identified a ‘de novo’ approach for exploring the relevance of the translation to the target population. The Arabic translation, for example, used open-ended interviews to find ‘the concept of oral health-related quality of life... [and] whether all ... the items... characterize... [oral health-related quality of life] in the new cultural environment’ (29). The Vietnamese version emerged in part from a committee of 10 lay persons to maintain the conceptual context of its origins; however, none of the translators proposed an alternative construct to the sick role theory underlying the original OHIP.

Translators in some cultures replaced or deleted items that confused respondents, pertaining mostly to psychological distinctiveness (self-consciousness or embarrassment), feelings (being upset; depressed; embarrassed or irritable) and activities (concentrating; going out), which either decreased (33) or increased (30–32, 34, 35) the number of items in the translation. There are, for example, two short forms of the Cantonese translation, one from a regression analysis of data collected during focus groups for the long form and the other from a translation of the original OHIP-14 (34). Together, they provide 19 unique items in Cantonese although the authors recommend using the original 14 items in translation for international comparisons. Similarly, the German short form has three versions derived from factor analyses of interviews for the long form modified by three different interpretations of items for a German population (31). The Hungarian (35), Japanese (32) and Malaysian (33) translations were all expanded in similar manner. The Korean OHIP (36) was translated uniquely without the statement ‘because of problems with your teeth, mouth or dentures’ that qualifies nearly three-quarters (71%) the original OHIP items.

Discussion
There are several different approaches for translating HRQoL instruments, and the 10 methodological papers collectively offer eight specific steps. All but seven of the OHIP translations followed at least five of the steps, even though, like other HRQoL instruments (3, 4), the reports on many of the steps lacked essential information for judging the quality or strength of the cultural equivalence. For example, all the reports on response options lacked detail beyond a simple reference to the 5-point Likert scale. Although forward–backward translations are the mainstays for linguistic equivalence, one-third of the translations did not use it, or at least the authors did not refer to it.
Table 2. Distribution of assessment steps and other details associated with cultural equivalence in translations of the Oral Health Impact Profile (OHIP)

<table>
<thead>
<tr>
<th>Translation (reference)</th>
<th>Number of questions in the translated OHIP</th>
<th>Long</th>
<th>Short</th>
<th>Steps for assessing cultural equivalence</th>
<th>Reference source of the translational method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albanian (53)</td>
<td>49</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(2)</td>
</tr>
<tr>
<td>Arabic–Saudi Arabian (29)</td>
<td>49</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(2)</td>
</tr>
<tr>
<td>Arabic–Lebanese (54)</td>
<td>14</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>(2)</td>
</tr>
<tr>
<td>Brazilian–Portuguese (21)</td>
<td>49</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(2)</td>
</tr>
<tr>
<td>Burmese (56)</td>
<td>14</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(2, 7)</td>
</tr>
<tr>
<td>Chinese–Cantonese (34)</td>
<td>49</td>
<td>✓</td>
<td>14, 19</td>
<td>✓</td>
<td>(26)</td>
</tr>
<tr>
<td>Chinese–Mandarin (51)</td>
<td>49</td>
<td>✓</td>
<td>14</td>
<td>✓</td>
<td>(26)</td>
</tr>
<tr>
<td>Chinese–Taiwanese (52)</td>
<td>49</td>
<td>✓</td>
<td>14</td>
<td>✓</td>
<td>(26)</td>
</tr>
<tr>
<td>Croatian (57, 58)</td>
<td>49</td>
<td>✓</td>
<td>14</td>
<td>✓</td>
<td>(2)</td>
</tr>
<tr>
<td>Dutch (22)</td>
<td>49</td>
<td>✓</td>
<td>14</td>
<td>✓</td>
<td>(2)</td>
</tr>
<tr>
<td>French–Canadian (15)</td>
<td>49</td>
<td>✓</td>
<td>14</td>
<td>✓</td>
<td>(2)</td>
</tr>
<tr>
<td>Georgian (59)</td>
<td>14</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(2)</td>
</tr>
<tr>
<td>German (30, 31)</td>
<td>53</td>
<td>✓</td>
<td>5–21b</td>
<td>✓</td>
<td>(2)</td>
</tr>
<tr>
<td>Greek (60)</td>
<td>14</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(2)</td>
</tr>
<tr>
<td>Hebrew (61)</td>
<td>14</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(2)</td>
</tr>
<tr>
<td>Hungarian (35)</td>
<td>49/5</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(2)</td>
</tr>
<tr>
<td>Italian (62)</td>
<td>14</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(3)</td>
</tr>
<tr>
<td>Japanese (32)</td>
<td>49/5</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(26)</td>
</tr>
<tr>
<td>Japanese (63)</td>
<td>14</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>(26)</td>
</tr>
<tr>
<td>Translation (reference)</td>
<td>Number of questions in the translated OHIP</td>
<td>Steps for assessing cultural equivalence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------------------------------</td>
<td>----------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Korean (36)</td>
<td>49</td>
<td>14</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Latvian (64)</td>
<td>49</td>
<td>14</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Macedonian (65)</td>
<td>49</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Malaysian (33)</td>
<td>45</td>
<td>14</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Maltese (66)</td>
<td>14</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Persian (67)</td>
<td>14</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Romanian (69)</td>
<td>14</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Russian (64)</td>
<td>49</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Sinhalese (71)</td>
<td>14</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Slovenian (22, 58)</td>
<td>49</td>
<td>14</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Spanish (72)</td>
<td>14</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Spanish</td>
<td>49</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Spanish-Chilean (73)</td>
<td>49</td>
<td>14</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Spanish-Chilean (74)</td>
<td>49</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Swedish (75)</td>
<td>49</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Turkish (76)</td>
<td>14</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Vietnamese (77)</td>
<td>14</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Total Number (%)</td>
<td>22 (61)</td>
<td>24 (67)</td>
<td>23 (64)</td>
<td>25 (69)</td>
<td>19 (53)</td>
</tr>
</tbody>
</table>

_de novo:_ personal interviews or focus groups with lay people to identify personal oral health-related concerns and experiences.

^aThesis in Portuguese.

^bThree versions OHIP-G5; OHIP-G14 & OHIP-G21.

^cAdolescents 12–21 years.
The systematic search, like all others, was limited by the comprehensiveness of the databases, the appropriate assignment of key words and the undetected bias of the investigators (19, 23). We did not measure with a Kappa score the agreement between the two investigators applying the inclusion/exclusion criteria to abstracts or full texts. The investigators applied the inclusion/exclusion criteria independently before achieving agreement by discussion. Nonetheless, the possibility remains of a combined selection bias, although its significance is unlikely to have much influence on our findings given the range of methodological and translational papers selected, and the likelihood of excellent agreement between two reviewers applying straightforward (present/absent) assessment criteria and a robust reconciliation process (37).

Our aim was to identify methods for translating psychometric instruments not to validate translations. Neither did we weigh or otherwise assign relative significance to each translational step, or assess the quality and effects of the translations. Perhaps item and construct equivalence are more important than other steps to cross-cultural equivalence, but we found no guidance on how to weigh or judge each step appropriately. Epstein et al. (38) believe that translations of psychometric items are prone to errors in style, intensity, frequency/time, breath and meaning, and recommended that the response options (interval scales) would benefit from the insights of expert multidisciplinary committees. Translations of interval scales are complicated further by a ‘reference group’ (39) or ‘relative deprivation’ (40) phenomena by which respondents tend to assess themselves relative to others. Consequently, requests for personal disclosures about self-consciousness, depression, embarrassment or irritability can generate responses biased towards social approval (41). Agarwal (42) suggests that the iterative process for verifying items for ‘construct validity’ is complicated and time-consuming. We found no evidence that this subtlety was explored for any of the 36 OHIP translations even when de novo methods were used. Tests of discrimination and convergence dominated claims of validity for the translations; however, Guillemin et al. (3) questioned their relevance to cultural equivalence. Others (2, 26, 42) believe that tests of these psychometric properties, including responsiveness, illustrate cultural acceptability of translations but only when all the measurements are reasonably reliable under the same test conditions and in both cultures.

Beyond references to the interviews and focus groups of the de novo method, there were few details on how the construct of the OHIP was explored for cultural relevance. The theory of sick role underlying the OHIP is challenged by the concept of people coping effectively with the adversities of life and ageing (13, 28, 43, 44). This concept of coping is overlooked by the OHIP as it focuses on the personal experiences of mouth problems without attending to how respondents manage their problems. Nutall et al. (45), for example, revised the OHIP construct from responses to the OHIP short form in the UK and Australia by emphasizing links between impairment and disability, and between functional limitation, pain and handicap, but without exploring how respondents coped with their disabilities or handicaps. It is surprising in tests of the OHIP translation that no details emerged from the de novo interviews or focus groups about these behavioural and psychosocial constructs. Possibly, the translators were focused so keenly on vernacular or linguistic translations of the OHIP items that they were unprepared to detect other conceptual relationships between health and illness (46). In contrast, vernacular equivalence received relatively careful attention leading to the addition, deletion or change of items from the Cantonese (34), German (30), Hungarian (35), Japanese (32) and Malaysian (33) translations. Clearly, this occurred to accommodate the nuances of language between cultures, particularly of how people see and explain themselves relative to others (16, 47–49).

If the minimal requirement for a translation is an accurate vernacular translation and reliability of psychometrics within a specific cultures, then most of the 36 OHIP translations are acceptable. But if the aim includes conceptual equivalence across cultural boundaries, then it is unlikely that more than half of them are acceptable (2), in large part because constructs such as quality of life are vague and difficult to measure (41). Indeed, each of the OHIP translations might be suitable for profiling oral health problems within one culture, but comparisons of scores between cultures are less useful, and especially without relevant cut-offs with which to interpret the scores (50).

**Conclusions**

An acceptable method for translating HRQoL instruments for cross-cultural equivalence has...
eight procedural steps, and most of the 36 OHIP translations involved at least five of them. Only translations to Saudi Arabian Arabic (29), Chinese Mandarin (51), German (40, 41) and Japanese (32) used all of the eight steps to claim cultural equivalence with the original OHIP.

References


63. Cross-cultural equivalence of the OHIP. Cross-cultural equivalence of the OHIP. Cross-cultural equivalence of the OHIP. Cross-cultural equivalence of the OHIP.