

## Advancing cross-cultural research on quality of life: Observations drawn from the WHOQOL development

Suzanne M. Skevington

*Department of psychology, University of Bath, Bath BA2 7AY, UK (E-mail: s.m.skevington@bath.ac.uk)*

Accepted in revised form 30 November 2001

### Abstract

This paper is concerned with methods and theory recently used in cross-cultural research. It begins by looking at why we need cross-cultural measures and why we need more of them. A discussion of the translation styles, and the issues underpinning their design is central to this discussion. Through a description of the World Health Organisation Quality of Life Assessment instruments and their development, the paper demonstrates how some of the earlier theoretical and methodological challenges have been addressed, together with a critique of its limitations.

**Key words:** Cross-cultural, Measurement, Quality of life, The World Health Organisation Quality of Life Assessment, Theory

### Definitions of quality of life

There are many definitions of quality of life (QoL); some writers have ventured that as many as one for every study. QoL is often confused (and not just by lay people) with standard of living, and a consideration of cross-cultural issues may assist us in clarifying the concept. Although global indicators like safe water in the home or within 15 min walking distance, have become standard of living goals for developing countries wanting to measure their progress in health care [1], some cross-cultural examples enable us to see that standard of living is neither synonymous with, nor a substitute for QoL. While there is a basic human need for adequate space to live in for instance [2], the number of square feet of living space per person in Los Angeles and Tokyo would be quite inaccurate indicators if used as proxies for QoL in these two countries. Standard of living tends to be assessed by a more 'objective' set of parameters than QoL, whereas QoL is about subjective perceptions of important aspects of a person's life that may or may not coincide with accepted consensus about

standard of living indicators. The two are relatively different but the relationship between them deserves further investigation.

QoL is about the meaning people derive about from the important aspects of their life; it is a social construction. Two hypothetical people who both feel identical pain intensity from a torn ligament in the lumbar spine resulting from a weekend gardening will not interpret the effect of this pain on their QoL in the same way because of their respective histories and experience. One person may largely ignore it and carry on with life as though the pain was unimportant but for the other, the pain will be a disabling disaster, with chronic depression and loss of time from work. Other life experiences have a range of highly individual interpretations.

The World Health Organisation Quality of Life Assessment (WHOQOL) group proposed that these perceptions and interpretations are rooted in that person's culture and this may be the first definition of QoL that has directly and formally incorporated cultural components as integral to its definition rather than acknowledging cultural

influence as an extraneous variable. The WHO-QOL Group defines QoL as an individual's perceptions of their position in life, in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept affected in a complex way by the person's physical health, psychological state, level of independence, social relationships, their relationship to the salient features of their environment [3, 4] and their spiritual, religious and personal beliefs. This theoretical orientation towards culture as a focal part of QoL understandings and expression has explicitly directed the methodology used in the WHOQOL project. At the same time, QoL is viewed as a rich interplay between multiple dimensions or domains that are perceived to be important and not as a uni-dimensional phenomenon.

#### **Why do we need cross-cultural measures of QOL?**

Perhaps the most urgent reason for designing cross-cultural measures is to serve evidence based medicine in the systematic monitoring of outcomes from multinational clinical trials. This information is needed to investigate those interventions that have not so far been routinely assessed using standardised measures; surgery, physiotherapy procedures, counselling, alternative therapies, medicines etc. The global value of many common treatments for chronic diseases e.g. the use of antidepressants, has still only been partially evaluated where it has been evaluated at all, (although this is changing) and often only in single cultural settings. Changes to QoL arising from events such as the reorganisation of health care and new ways of delivering treatment are also of great importance in a patient-centred system. Such measures are most useful where they have been developed using culture-sensitive methods.

Secondly, cross-cultural measures enable comparisons to be made about QoL in different cultural or social groups and between those receiving care in different health care settings. This is valuable information for policy-makers and planners at regional, national and international levels. It could be particularly useful in the HIV pandemic, for instance, where so many different cultural

groups require QoL assessment at the time of voluntary HIV testing and counselling (VCT), and afterwards.

At the same time such pragmatic cross-cultural studies also provide theoretical insights into whether QoL is a universal or relativist concept. In a recent review of the area, Bullinger [5] observed that the dimensions included in current generic as well as specific scales show an 'astonishing degree of correspondence across instruments'. Where individual open-ended statements were made about QoL, these also revealed very few differences either between cultures or between individuals. This view is supported by data collected by the IQOLA group [6] where SF-36 profiles from four European countries have shown a very high degree of similarity. Results from the WHOQOL Group [4, 7–9] with contributions from many countries world wide also support this position. However researchers working with an idiographic approach like the SEIQOL that seeks to understand the individual as unique, rather than as another point on a group dimension [10], dissent from this view [5]. They take the view that the many individual differences between people are more important than their similarities. This debate may be partly explained by the choice of methodology and partly by the choice of statistical techniques used to analyse pooled (group) data [11].

#### **Why do we need more cross-cultural measures?**

There are now a number of QoL instruments that have been well developed for generic use and the need to create cross-cultural applications was a key aim for developers, from early in the history of QoL measurement e.g. the sickness impact profile (SIP) [12]. By 1996, six instruments were identified that were available in six or more European languages [13], so it is legitimate to ask whether the field needs to develop more scales. But methods of translating instruments have not changed much since the SIP was published and researchers still remark on widely shared translation difficulties that appear to have no solution ([14] on the Nottingham Health Profile). The IQOLA group recorded problems with trying to make their language versions more culturally appropriate in an account of the sequential translation procedure

used on the SF-36 in the conversion from the original US English version into the majority languages of 10 countries [15]. However an advance was made through the use of a parallel approach adopted by the European Organisation for Research and Treatment of Cancer Quality of Life Working Group (EORTC), in the design of the EORTC-QLQ-C30 [16], where the new scale was assembled by conceptually grounding it in existing scales from different countries [13].

Furthermore inspection of comparative data reveals that when the psychometric properties of a source instrument are compared with those of a new translation from a rigorously tested instrument in a target culture, the accepted parameters of reliability and validity tend to be notably poorer for the target culture than they were for the source. For example, the Spanish version of the MOS-HIV [17] and the Xhosa translation of the SF-36 [18] both showed poorer psychometric performance and across a wide range of psychometric criteria, when compared to the US version. In this way, the properties of subsequent versions from target cultures are inevitably compared with those from the primary source. Together this body of evidence suggests that there are still barriers that need to be overcome when developing instruments in different languages. This seems especially urgent when reviews of the literature also reveal that where generic profiles are concerned, the large majority have been developed in North America [19]. This means that a US-centred view of QoL inevitably pervades the form and content of the most popular scales in use and their results.

Before discussing translation styles, it is important to briefly outline Pike's [20] classical work on emics and etics that underpins this thinking. The emic approach aims to uncover and describe the specific behaviours of a particular culture through using the cultures own terms. Emics are features of a phenomenon – in this case QoL – that only apply to a particular society. Arising from the ethnographic methods of cultural anthropology, those working in this approach aim to understand the culture they are interested in through the eyes of its community. Focussing on the uniqueness of cultures (rather than individuals), it bears some similarities to the idiographic approach. Etics in contrast, are ostensibly culture-free, general or universal aspects of QoL. However it is question-

able whether researchers are entirely able to shed the profound influences of their own cultural background when deriving etics and consequently an imposed etics or pseudoetics approach seems a more realistic description [21]. In practice, the research process typically starts with an imposed etics approach and then the concepts, tests etc. are modified to match the emic viewpoint. One of the hallmarks of a successful investigation is that the process concludes with a modified or derived etic approach that enables valid comparisons to be made between language versions [22].

Four broad types of translation procedure have been used in QoL research [23], although there are some variations on these themes. An ethnocentric approach like that taken with the SIP, implicitly assumes that it is possible to translate an instrument developed in one culture into a second language without attending to how appropriate the instrument will be to that target culture. This approach suffers from the questionable assumptions that these concepts are transferable, that the two languages are so closely related that translation is meaningful and that the method of questioning is as acceptable to the target as it is to the source. A distortion of the results obtained from assessing partially or wholly culturally inappropriate health constructs and concomitant difficulties in interpreting the resulting data, are two of the principal disadvantages to this approach. A second, more pragmatic approach seeks out the common conceptual ground between two cultures and measures these issues; a variation of this approach appears to have been used by the EORTC. Although greater conceptual equivalence is achieved with this method, the requirement of a comprehensive construct of QoL cannot be entirely satisfied.

The third 'emic plus etic' approach has been applied, where there is some level of overlap between the target and source. Broadly speaking, the common ground between the two cultures is addressed and translated (etic) and then at another stage, issues specific to the culture (emic) are targeted separately. This approach had been used in the development of the WHOQOL where the international core instrument contains derived etics that enable comparisons between versions to be made. Additional national items (emics) were translated, evaluated and included at a later stage [11]. In the fourth case, items from a source

instrument are interpreted for use by the target culture without presenting the instrument to members of the target for assessment and there are plentiful examples of specific and generic scales that have applied the latter method in cross-cultural work [24].

These approaches may result in limited semantic, conceptual and technical equivalence between source and target instruments [23] and this is the crux of the translation conundrum. Semantic equivalence occurs where words in other languages indicate what is being said in the first language or provide a sign for it. Their connotation or emotional meaning must also be the same if full equivalence is to be achieved, and the ability to capture these emotional nuances will depend on the proficiency and communication skills of the translator. Conceptual equivalence between instruments refers to the elaborate or restricted nature of a concept or dimension and the various meanings attributed to these different aspects. To achieve full conceptual equivalence, all the important aspects of different facets of QoL should be represented as items in an instrument. A Zulu dictionary published in 1946 shows that the English word for a cut has many words in Zulu where it refers to castration and circumcision, as well as to cutting the hair, skin, ears for rings, nose, forehead, a wound in the forehead and a gash [25]. This means that several questions on this area of QoL would need to be included in any Zulu instrument that included pain as one of its facets, whereas the British English version only seems to require two. However achieving conceptual equivalence brings headaches for those concerned with scoring, as different numbers of items in different language versions complicate the way that versions of the same instrument can be scored. In cross-cultural work, researchers seek assurance that their instruments (concepts, tests, appraisal process) hold a similar meaning (or lack of it), for all the groups that are being compared [26]. Translation equivalence helps to establish conceptual equivalence [27]. Technical equivalence is concerned with the feasibility of the nature and method of questioning. Direct questioning may prove to be difficult or even impossible for particular areas of concern in some cultures e.g. sexual activity in Britain, and in translation, some styles of questioning may be seen as rude, pa-

tronising, and so on. These barriers may be partly (but not entirely) overcome through a painstaking and repeated process of translation and back-translation using monolingual and bilingual groups until both researchers and linguists are satisfied. Together, all this in turn helps to generate the highly desired state of metric equivalence or the equivalence of results or data [26].

### **Changes in cross-cultural assessment**

In 1991 the WHOQOL Group through the World Health Organisation began a project that aimed to tackle some of the problems that beset cross-cultural measurement of QoL relating to health. At the start, health experts from many countries worldwide, generated and discussed several hundred aspects or facets of QoL, and provisionally aggregated them into clusters (domains). International agreement on 33 important facets of QoL in health and their definitions led to the compilation of a manual that also contained examples that had been written with lay people in mind. This could be used as guidance to focus groups convened to generate items for the WHOQOL. These small groups of patients, health professionals and well people from the community (including informal caregivers) also gave spontaneous definitions of QoL, rated the importance of the proposed facets and commented on any important omissions [4].

This evaluative scale is designed for sick and well younger adults (16–65 years). The WHOQOL was intended to provide an evidence-base for multinational clinical trials, permitting differences between various sick and well populations to be appreciated, and for the assessment of other population differences in health. Furthermore it was intended to create a subjective self-report assessment for group and individual use, should the properties prove to be good. Its use in the allocation of health resources by taking into account the patient's perspective and its integration into policy-making as part of a patient-centred approach to the delivery of health care could also not be ruled out.

A new translation and back-translation methodology using alternate multilingual and bilingual panels to review the instrument using an iterative process was developed and applied at each stage of

the project [23]. The working language of the group was English. High quality back-translations are the ‘acid test’ of a successful translation procedure and this quality control was available to centres from the coordinating group at WHO Geneva. The identification, operationalisation and translation of emic constructs were central to this activity. Focus groups around the world told us that spirituality, religion and personal beliefs were important to their QoL and health but this issue had not been included in the original framework, and so appropriate items were subsequently added as a sixth domain. Items were phrased by the focus groups using their language and concepts. This way, the WHOQOL was designed by the users for the users, and for this reason, employs a highly patient-centred approach in its development as an instrument.

The WHOQOL group is a collaboration of researchers, clinicians and scientists who have been working together from common, consensually agreed international protocols at every stage in the development of the project over 10 years. This contrasts with the modus vivendi of other QoL groups (the EORTC is a notable exception) where researchers from non-source cultures become members of the working group after the initial version of the instrument has been completed; usually at the time they seek to translate the instrument for use in their culture. This means that the cultural input of the newcomers to the instrument has been by-passed at the conceptual stage of instrument formulation. There were 15 countries worldwide, in the original WHOQOL collabora-

tion but the new methodology that created the manual (and other protocols) as a ‘blueprint’ can potentially facilitate the production of thousands of equivalent language versions. Today there are around 40 language versions. This high level of co-operation, and willingness to move forward simultaneously during development, has enabled advances to be made to the methodology with less of the accompanying frustration, disappointment and effort that the sole use of sequential translation incurs. It also removes the need to describe versions as target or source, as all versions were derived from the same common methodology that was agreed through the consensus of the collaboration. This ‘spoke-wheel’ methodology [28], where the procedure is akin to the spokes and hub of a bicycle wheel, not only improves equivalence at semantic, conceptual and technical levels but at the same time represents a ‘fast track’ to simultaneous instrument development in many languages [29].

The WHOQOL-100 was derived from pilot work on 235 items and the resulting scale contained 100 items, hierarchically organised into 25 facets (including a general facet on QoL and health) and originally grouped into six domains of QoL. Its properties have been rigorously tested on a large, heterogeneous, global sample of sick and well people ( $n = 4802$ ) using traditional psychometric techniques. Structural equation modelling has been used to reaffirm construct validity and indicates a better fit for a four domain structure [8] (see Table 1). More recently Rasch techniques have also been applied to investigate the structure [30].

**Table 1.** The WHOQOL structure: domains and facets

Physical	Psychological	Social relationships	Environment
Pain and discomfort	Positive feelings	Personal Relationships	Physical safety and security
Energy and fatigue	Thinking, learning, memory and concentration	Practical social support	Home environment
Sleep and rest	Self-esteem	Sex	Financial resources
Mobility	Body image and appearance		Health and social care: availability and quality
Activities of daily living	Negative feelings		Opportunities for acquiring new information and skills
Dependence on medication or treatment	Spirituality, religion and personal beliefs		Participation and opportunities for recreation and leisure
Working capacity			Physical environments
			Transport

The WHOQOL-100 and an extracted short form of 26 items for the WHOQOL-BREF are robust [8, 9, the WHOQOL Group, unpublished]. At each stage, items and facets have been excluded where the results were found to be inconsistent in more than half the centres contributing to the analysis, so that only issues reflecting a high level of global consensus were retained. In the process of creating a longer and shorter evaluative tool, item reduction was completed in several stages. During piloting, nearly 1000 items were selected from a pool of over 2000 using 11-item selection criteria. These criteria referred to conceptual issues, like whether the item assessed part of the facet definition, to technical issues connected with administration, like being amenable to a rating scale, shorter rather than longer, and to semantic issues, such as avoiding ambiguous phrases/words and double negatives. The reduced item pool was assessed by all centres and consensus used to select the best items for the international pilot version (325 items). To do this, conceptual item clusters were identified within facets e.g. for pain and discomfort, five clusters described the intensity/frequency of pain, coping with pain, pain control, distress from pain and disability caused by pain. Clusters were first ranked by each centre for importance to the QoL of that culture, then centres ranked questions within the clusters. Ratings of adequacy were also utilised for larger groupings of questions.

Statistical analyses were completed first for each centre and then for pooled data, on the items normative properties, internal consistency reliability contribution and correlation with other items, facets and domains. Non-significant and negative correlations, non-significant discrimination between sick and well samples, and overlap with other items (strong associations) were also used to exclude items. Stepwise multiple regression with other techniques extensively recorded by WHOQOL Group [8], determined the item that best explained the total score (excluding the general items), then the second best, and so on. As only one item per facet was added to the model at each stage, the selected item had least overlap with the other items included (A. Harper, pers. comm.). A few substitutions were made where items were too highly correlated with the non-parent domain. Domains were assessed for sick-well discrimina-

tion, good reliability etc, before the integrity of the entire scale was subjected to a full psychometric evaluation. A WHOQOL panel established whether items in each domain covered the concept outlined in the agreed definitions, and whether a cohesive and interpretable entity of QoL with good 'gestalt' and construct validity had been achieved (A. Harper, pers. comm.). In the light of these results and after the EQS model was applied, further minor adjustments were made to better explain the concept [8]. Using this reflexive, iterative and integrated qualitative and quantitative procedure, the conceptual features of QoL agreed by the group continued to be considered at the time that quantitative information on item selection was also under review.

The psychometric properties of the short-form WHOQOL-BREF have been tested on the original pilot data ( $n = 4802$ ), field data ( $n = 4104$ ) and data ( $n = 2369$ ) from new centres and the results are similar to the WHOQOL-100 and of a high calibre [9]. In the BREF, 95% of the total facet score variance was explained for the four domains (range 91–77%). Cronbach's  $\alpha$  for domains was acceptable except for the small, 3-item social relationships domain ( $\alpha = 0.66$ ). Domain discrimination between sick and well groups was good ( $p < 0.001$ ). Comparative fit indices of the four domain model using 24 facet items for the total was 0.90 for the original and new data [9] (Table 1).

Several other features of the WHOQOL are worthy of mention. Importance items relating to each facet were appended to the WHOQOL-100. These have assisted in the selection of those facets that were perceived to be internationally important or very important, although as Hagerty et al. [31] observe, these have not been used to weight facets in scoring as originally intended, for psychometric reasons. Recent work suggests that they may provide valuable information about QoL in their own right, and extended discussion of this issue is forthcoming (S.M. Skevington and K.A. O'Connell with the WHOQOL Group, unpublished).

Secondly, four new types of response scale were developed to accompany the nationally derived wording for items. This was carried out using a novel methodology whereby the WHOQOL group first agreed translatable labels that could be

inserted at the poles of the five-point scales. But the three words that were used at the 25, 50 and 75% intervals were derived through field-work in the centres, that required respondents to place labels on a 100 mm line, so that the most appropriate descriptor for each interval could be selected [32, 33]. Consequently, even in the English-speaking centres of Melbourne and Bath, the three intermediate labels are different for each response scale. This work therefore supports the view that response scale descriptors do not cross international boundaries but must be developed locally, so that the language is most appropriate to that culture. Like items, their wording will in time also become outdated.

Thirdly, to take account of the problems of conceptual equivalence, national items could be added to the highly comprehensive international core items (within defined limits) to round out the concept in that language and culture, where such issues were known to be important. For example, skin colour in India is central to people's QoL, whereas it would be of marginal importance to the majority of English respondents. The inclusion of these national items depended on whether they would be as robust in psychometric terms, as items from the international core. It was also important that they did not replicate the meaning of existing core items, and the complimentary techniques of multi-dimensional scaling and cluster analysis were used to examine the conceptual structure of facets containing national items in each centre, to ensure that this did not occur [11]. Not all centres proposed national items because their focus groups did not cite additional important issues. The English version has two national items on 'feeling fed up' and happiness.

Lastly, it was acknowledged that even a highly comprehensive generic instrument like the WHOQOL might not address certain special QoL issues that arise when people have a particular disease or condition. These could be developed in a similar way to the core instrument and be appended to it. For instance, for cancer patients in therapy, vomiting and nausea may be salient to their QoL but it is unlikely that these questions would be appropriate to the general population or even the majority of sick people, and would therefore be at best, redundant and at worst, an irritating extra burden to those completing the scale. Generating

specific items for use with a particular disease or condition follows the original WHOQOL protocol and must include three or more culturally diverse centres to be acceptable. The new WHOQOL-HIV instrument has been created in this way and includes data from a range of centres to test and reduce items generated by focus groups of people with HIV infection and well people (K.A. O'Connell and S.M. Skevington with the WHOQOL-HIV Group, unpublished). Although the 100-item version has the strength of being an accessible, feasible, highly comprehensive, valid and responsive tool [34] the inclusion of specific items may not only improve acceptability but should also increase the sensitivity of the instrument to changes in a patient's clinical condition.

### Critique

Like other instruments, the WHOQOL has its shortcomings and some of these have been highlighted in a recent review of general QoL measures [31]. Using a Delphi technique, the authors agreed 14 criteria that they believed to be important to QoL 'indexes' (sic) for use in public policy, and assessed 22 measures against each.

One criticism of the WHOQOL was that there was no justification or rationale for the choice of domains. Although not fully reported in early publications, the domain structure was derived from the context in which particular discourses were located in focus group discussions, as well as from an extensive review of the literature on QoL and the use of expert international opinion. Haggerty et al. [31] identify some issues as important omissions from the WHOQOL concept. Material well-being may well have a place in a health-related QoL measure but had this been important, then the focus groups would have had the opportunity to indicate this. Employment is addressed under working capacity and pilot work with 29 facets showed that a second facet on work satisfaction did not perform well across cultures and so it was removed from the final structure of the WHOQOL-100 [7]. The WHOQOL facet on spirituality, religion and personal beliefs originally held domain status and this seems to have been overlooked by these reviewers. Its inclusion is a radical departure for generic scales where few

designers have attempted to operationalise this neglected area. During the WHOQOL development it was recognised that the conceptualisation and measurement of spirituality, religion and personal beliefs was limited, but for pragmatic reasons it was not possible at that time to develop it. This agenda is currently being revisited by the WHOQOL Group in a program of international work with centres representing the widest range of belief systems, to develop an assessment of spirituality, religion and personal beliefs in relation to QoL and health.

The WHOQOL work is said to hold a 'deficit view' of the construct on QoL. Hagerty et al. [31] say that through this perspective, full health is equated to maximum QoL and QoL can be discounted through ill-health. But the definition of QoL derived by the WHOQOL group implies that QoL is about the many and varied meanings that people bring to important aspects of their lives and this is a much more complex analysis than assumed. It is this very feature that the WHOQOL Group have tried to challenge. The detailed inclusion of many diverse symptoms in the WHOQOL would have driven this health-related QoL scale in the very direction that Hagerty et al. erroneously conclude, and this was one of the reasons why it was decided to avoid this orientation. As mentioned earlier, counting symptoms or assessing symptom intensity is not integral to the WHOQOL because it is not possible to assume that people with serious breathlessness, nausea or insomnia will necessarily report a poor QoL.

Hagerty et al. [31] comment on the absence of 'objective' items in the WHOQOL assessment. These types of question were considered for inclusion in the scale and rejected after preliminary pilot work [3]. But it was recognised that many QoL scales contain 'objective' items and that a subjective self-report measure was needed that could be used alongside existing measures where necessary. The WHOQOL group takes the view that QoL is about the meaning of different aspects of life for individuals. As it is not possible to observe meanings because they are internal states then QoL must necessarily be measured subjectively. Objective measures are of course important but perhaps this perspective should no longer be described as QoL, to avoid labelling confusion.

The WHOQOL has also been criticised for combining different response scales in the same scoring unit, for changes to the factor structure in successive publications with different samples, for applying a dual conceptual and statistical approach where others would have used only a statistical rationale (despite the meaning of the solution) and for not declaring data reduction processes. These issues are now further addressed within this paper.

Comprehensiveness can come at a price. In a few cases, the wording of two items from the same facet reflect subtle linguistic or semantic differences that may be difficult to fully appreciate during rapid completion or operationalise in some translations, so duplication may be perceived. A shorter, more succinct version of the WHOQOL-100 may need to be derived in the future. While the focus group procedure went some way to obtaining questions that were meaningful and relevant to the users, full cognitive debriefing procedures might have enhanced this process.

Lastly, while some researchers might view the WHOQOL Groups procedure as too laborious and complex for regular instrument development, it could be argued that the considerable investment required to create a high quality cross-cultural instrument demands this level of fastidious attention to detail. Given the number of generic QoL instruments in existence, perhaps such projects should not be undertaken in the future unless these procedures can be bettered.

## Conclusion

Ending on a theoretical note, the WHOQOL project has employed elements of etic and emic perspectives and its features will be re-evaluated with reference to Berry's [35] four criteria needed to distinguish these approaches. In the WHOQOL case, QoL is studied both from within the system through the centres (emic), and from an outsider's point of view from the other perspectives in the international group (etic). Through its conceptual development and analytical procedures, the findings of each centre are addressed and appraised culture by culture (emic) but QoL is also examined in many cultures so enabling some comparisons to be made (etic). Furthermore, the structure of QoL

in the WHOQOL was discovered using a bottom-up approach with focus groups to find out what dimensions were salient and consensually agreed as important across cultures (emic) but also top-down, using expert opinion and literature review (etic). The WHOQOL approach enables more than just a relativist, comparative, multiple cross-national exercise to be undertaken, but essentially employs a universalist approach to the interpretation of QoL [21]. Its universality arises from the simultaneous development of the instrument in many different cultural settings, a common set of agreed facet definitions and a common methodology for question development and selection. Additionally it contains a global pool of questions with which any national items have to compete in psychometric and conceptual terms [8]. Its dimensions have been found to be reliable and valid across many different centres [7–9, the WHOQOL Group, unpublished] and structural equation modelling together with other statistical tests show that there is a substantial level of agreement across 15 countries about the multiple dimensions and structure of the concept [36].

### Acknowledgements

An earlier draft of the paper was prepared for a meeting of the ESRC Multidisciplinary group on Health-related Quality of Life and its use in Decision-Making. The author acknowledges discussion with the group and thought-provoking comments by anonymous referees.

The paper is based on experience obtained as part of the WHO study to develop a quality of life measure (WHOQOL). The collaborators in this study have been at WHO Geneva: Dr J. Orley and Dr Rex Billington assisted by Dr Willem Kuyken, Dr Norman Sartorius and Dr Mick Power. In the Field Research Centres, collaborating investigators are Prof Helen Herrman, Dr H. Schofield and Ms B. Murphy, Univ. of Melbourne, Australia, Prof Z. Metelko, Prof S. Szabo and Mrs M. Pibernik-Okanovic, Institute of Diabetes, Endocrinology and Metabolic Diseases and Dept. of Psychology, Faculty of Philosophy, Univ. of Zagreb, Croatia, Dr N. Quemada and Dr A. Caria, INSERM, Paris, France, Dr S. Rajkumar and Mrs Shuba Kumar, Madras Medical College, India, Dr S. Saxena, All

India Institute of Medical Sciences, Delhi, India, Dr D. Bar-On and Dr M. Amir, Ben Gurion Univ., Beer Sheeva Israel, Dr Miyako Tazaki, Dept. of Science, Science Univ. of Tokyo, Japan and Dr Ariko Noji, Dept. of Community Health Nursing, St. Lukes College of Nursing, Japan, Dr G. van Heck and Mrs J. de Vries, Tilburg Univ., The Netherlands, Prof J. Arroyo-Sucre and Prof Pichard-Ami, Univ. of Panama, Panama, Prof M. Kabanov, Dr A. Lomachenkov, and Dr G. Burkovsky, Bekhterev Psychoneurological Institute, St. Petersburg, Russia, Dr R. Lucas Carrasco, Barcelona, Spain, Dr Yooth Bodharamik and Mr Kitikorn Meesapaya, Institute of Mental Health, Bangkok, Thailand, Dr D. Patrick, Ms M. Martin and Ms D. Wild, Univ. of Washington, Seattle, USA and Prof W. Acuda and Dr J. Mutambirwa, Univ. of Zimbabwe, Harare, Zimbabwe. An international panel of consultants includes: Dr N.K. Aaronson, Dr P. Bech, Dr M. Bullinger, Dr Henian Chen, Dr J. Fox-Rushby, Dr C. Moinpur and Dr R. Rosser. Consultants who have advised WHO at various stages of the development of the project have included: Dr D. Buesching, Dr D. Bucquet, Dr L.W. Chambers, Dr B. Jambon, Dr C.D. Jenkinson, Dr D. De Leo, Dr L. Fallowfield, Dr P. Gerin, Dr P. Graham, Dr O. Gureje, Dr K. Kalumba, Dr Kerr-Corea, Dr C. Mercier, Mr J. Oliver, Dr Y.H. Poortinga, Dr R. Trotter and Dr F. van Dam.

### References

1. World Health Organisation. World Health Statistics Annual. WHO: Geneva, 1987.
2. Doyle L, Gough IR. A Theory of Human Need. London: Macmillan, 1991.
3. The WHOQOL Group. Study protocol for the World Health Organisation project to develop a quality of life assessment instrument (the WHOQOL). *Qual Life Res* 1993; 2: 153–159.
4. The WHOQOL Group. The development of the World Health Organisation Quality of Life Assessment instrument (the WHOQOL). In: Orley J, Kuyken W (eds), *Quality of Life Assessment: International Perspectives*. Berlin: Springer-Verlag, 1994; pp. 41–60.
5. Bullinger M. Cognitive theories and individual quality of life assessment. In: Joyce CB, O'Boyle CA, McGee H (eds), *Individual Quality of Life: Approaches to Conceptualisation and Assessment*, Amsterdam: Harwood Academic Publishers, 1999, pp. 29–40.

6. Aaronson NK, Acquadro C, Alonso J, et al. International Quality of Life Assessment (IQOLA) Project. *Qual Life Res* 1992; 1: 349–351.
7. The WHOQOL Group. The World Health Organisation Quality of Life assessment (WHOQOL): Position paper from the World Health Organisation. *Soc Sci Med* 1995; 41(10): 1403–1409.
8. The WHOQOL Group. The World Health Organization Quality of Life Assessment (WHOQOL): Development and general psychometric properties. *Soc Sci Med* 1998; 46(12): 1569–1585.
9. The WHOQOL Group. Development of the WHOQOL-BREF Quality of Life Assessment. *Psychol Med* 1999; 28: 551–558.
10. Cook M. *Levels of Personality*, 2nd ed. London: Cassell, 1993.
11. Skevington SM, Bradshaw J, Saxena S. Selecting national items for the WHOQOL: Conceptual and psychometric considerations. *Soc Sci Med* 1999; 48: 473–487.
12. Bergner M, Bobbitt RA, Carter WB, et al. The Sickness Impact profile: Development and final revision of a health status measure. *Med Care* 1981; 19: 787–805.
13. Bullinger M. The challenge of cross-cultural quality of life assessment. *Psychol Health* 1997; 12: 813–824.
14. Hunt SM. The researchers tale: a story of virtue lost and regained. In: Joyce CB, O'Boyle CA, McGee H (eds), *Individual Quality of Life: Approaches to Conceptualisation and Assessment*, Amsterdam: Harwood Academic Publishers, 1999; pp. 225–232.
15. Wagner AK, Gandek B, Aaronson NK, et al. Cross-cultural applications of the content of the SF-36 translations across 10 countries: Results from the IQOLA project. *J Clin Epidemiol* 1998; 51(191): 925–932.
16. Aaronson NK, Cull A, Kaasa S, et al. Language and translation issues. In Spilker B (ed.), *Quality of Life and Pharmacoeconomics in Clinical Trials*, 2nd ed. Philadelphia: Lippincott-Raven, 1996; pp. 575–587.
17. Badia X, Podzamczar D, Garcia M, LopezLavid C, Consiglio E. A randomised study comparing instruments for measuring health-related QoL in HIV-infected patients. *AIDS* 1999; 13: 1727–1735.
18. O'Keefe EA, Wood R. The impact of human immunodeficiency virus (HIV) infection on quality of life in a multi racial South African population. *Qual Life Res* 1996; 5: 275–280.
19. Fox-Rushby J. The relationship between health economics and health related quality of life. In: Orley J, Kuyken W (eds), *Quality of Life Assessment: International Perspectives*. Berlin: Springer-Verlag, 1994; pp. 61–74.
20. Pike KL. *Language in Relation to a Unified Theory of the Structure of Human Behaviour*. The Hague: Mouton, 1966.
21. Berry JW, Poortinga YH, Segall MH, Dasen PR. *Cross-Cultural Psychology: Research and Applications*. Cambridge: Cambridge University Press, 1992.
22. Berry JW. Imposed etics-emics-derived etics: The operationalisation of a compelling idea. *Int J Psychol* 1989; 24: 721–735.
23. Sartorius N, Kuyken W. Translation of health status instruments. In: Orley J, Kuyken W (eds), *Quality of Life Assessment: International Perspectives*, Berlin: Springer-Verlag, 1994; pp. 3–19.
24. Bowling A. *Measuring Disease: A Review of Disease-Specific Quality of Life Measurement Scales*. Buckingham: Open University Press, 1995.
25. Skevington SM. *Psychology of Pain*. Chichester: Wiley, 1995.
26. Berry JW, Dasen PR. *Culture and Cognition*. London: Methuen, 1974.
27. Frijda N, Jahoda G. On the scope and methods of cross-cultural research. *Int J Psychol* 1966; 1: 109–127.
28. Skevington SM. Investigating the relationship between pain and discomfort and quality of life, using the WHOQOL. *Pain* 1998; 76: 395–406.
29. Anderson TRT, Aaronson NK, Wilkin DT. Critical review of the international assessments of health-related quality of life. *Qual Life Res* 1993; 2: 369–395.
30. Lepage A, Ecosse E, Andrich D and the WHOQOL Rasch Scientific Committee. Methodological issues in using the Rasch model to develop a quality of life index: The analysis of four WHOQOL-100 data sets (Argentina, France, Hong Kong, UK). Paper to 5th Annual meeting of the International Society for Quality of Life Research, Baltimore, USA, 1998.
31. Hagerty MR, Cummins RA, Ferriss, AL, et al. Quality of life indexes for national policy: review and agenda for research. *Soc Indicators Res* 2001; 55: 1–96.
32. Szabo S, Orley J, Saxena S on behalf of the WHOQOL Group. An approach to response scale development for cross-cultural questionnaires. *Eur Psychol* 1997; 2(3): 270–276.
33. Skevington SM, Tucker C. Designing response scales for cross-cultural use: Data and experience from the development of the UK WHOQOL. *Br J Med Psychol*, 1999; 72: 51–61.
34. Skevington SM, Wright A. Changes in the quality of life of patients receiving anti-depressant medication in primary care: Validation of the WHOQOL-100. *Br J Psychiatry* 2001; 178: 261–267.
35. Berry JW. On cross-cultural comparability. *Int J Psychol*, 1969; 4: 119–128.
36. Power M, Bullinger M, Harper A and the WHOQOL Group. The World Health Organisation WHOQOL-100: Tests of the universality of quality of life in 15 different cultural groups world-wide. *Health Psychol* 1999; 18(5): 495–505.
37. Bullinger M. Ensuring international equivalence of quality of life measures: problems and approaches to solutions. In: Orley J, Kuyken W (eds), *Quality of Life Assessment: International Perspectives*, Berlin: Springer-Verlag, 1994; pp. 33–40.

*Address for correspondence:* Suzanne M. Skevington, Department of psychology, University of Bath, Bath BA27AY, UK  
 Phone: +44-1225-826830; Fax: +44-1225-826752  
 E-mail: s.m.skevington@bath.ac.uk