

New approach to translating instruments for cross-cultural research: a combined qualitative and quantitative approach for translation and consensus generation

ATHULA SUMATHIPALA, Section of Epidemiology and General Practice, Institute of Psychiatry, Kings College and Bethlem and Maudsley NHS Trust, London

JOANNA MURRAY, Section of Epidemiology and General Practice, Institute of Psychiatry, Kings College, London

ABSTRACT *Translation is the very essence of ethnographic research. The current practice of translation and back translation of an instrument is considered by some as inadequate. We used a combined qualitative and quantitative method: a nominal group to translate and assess the extent of agreement (consensus measurement) on the appropriateness of the translation and resolve disagreement (consensus development).*

A panel (n = 9) fluent in English and Sinhala, translated the 21-item Bradford Somatic Inventory, independently of one another. Each member rated each individual translation independently, for conceptual and semantic equivalence based on predefined quantitative measures for consensus. Translations failing to reach consensus were discussed by the group for modifications and subjected to the process of consensus generation. Translations agreed to be inappropriate were excluded. Translations agreed as appropriate were subjected to ranking and the final selection.

The process of translation confirmed that only part of the original meaning could be expressed in local terms. This resulted in the loss of some connotations making the item too narrow, or conversely in having to expand the original meaning in the local language terms making the translated item broader.

No single participant produced a significantly different contribution to the final outcome than other participants. The range of contribution to the finally selected questionnaire was five to eight translations with median of seven. Therefore the likelihood of one individual producing a translation of an instrument with 100% probability of acceptance to a wider group is unrealistic.

A group is better placed to translate, modify or eliminate inadequate or ambiguous items and generate culturally appropriate translation with semantic and conceptual equivalents. The process we describe, to assess and guide translations by a group, is better than arbitrary decision making by one or more individuals handling the translations and back translations. The combined qualitative and quantitative approach provides a more rigorous, systematic and contextual approach to translations.

Introduction

Cross-cultural studies can be approached from two different perspectives, which, together, have been called the 'etic-emic' paradigm (Brislin, Lonner and Thorndike, 1973). The 'emic' perspective involves the evaluation of the studied phenomenon from within a particular culture and context, aiming to understand its significance and relationship with other intracultural elements. Although a thorough understanding of concepts relevant to one culture is obtained using this approach, they are not necessarily comparable to those of other cultures. The 'etic' perspective, on the other hand, involves the evaluation of a phenomenon from

outside a particular culture, aiming to identify and compare similar phenomena across different cultural contexts (Bravo, Woodbury-Farina, Canino, Rubio-Stipec, 1993). Therefore, the use of standardized instruments, including structured interviews, pencil-and-paper-tests and questionnaires, developed outside a particular culture is essential for comparing and understanding phenomena across linguistically and culturally different populations.

According to Kleinman (1987), translation is the very essence of ethnographic research. Instruments are developed in a vernacular that may be quite difficult to translate into another language and strict lexical

translations are often meaningless in non-Western cultures. He criticized medical research, including psychiatric research, for often proceeding as if translations were a nuisance to be quickly handled.

Essential considerations in translating instruments

There are five kinds of validity problems when working with translations (Gaviria, Pathak, Michell, Flaherty, Winthrop, Martinez, Pacheco, Richman, Birz, 1984; Bravo et al., 1993):

- content validity – each item is to assess a content that is relevant to each culture under investigation;
- semantic validity – words used in the original and the translation must have a similar meaning;
- technical validity – a similar effect is to be achieved by the measuring technique in different cultures (this is relevant to societies with poor literacy rates);
- criterion validity – whether responses to similar items relate to the same normative concept in two cultures;
- conceptual validity – this requires that responses to an interview relate to a theoretical construct within the culture.

Current approaches

Current translation practices include back translation, decentring, the bilingual approach, and the committee approach (Drenann, Levett and Swarts, 1991; Guillmin, Bombardier and Beaton, 1993). Back translation (Drennan et al., 1991) is defined as translating while trying to change as little as possible in the final version of the original language. This procedure yields two versions of the instrument, one in source language, the original, and the back translated version. Comparison of these two versions provides a basis for further exploration of problematic areas and concepts.

Decentring (Drennan et al., 1991), on the other hand, does not require that the final version of the instrument is close in form or content to the original. It requires that the conceptual domain of the instrument, rather than the structure of any particular language, determine how the final product will look. This technique considers the source and the final version equally important (Werner and Campbell, 1970).

Kleinman (1987) views the handling of translations as technical, involving a process of translation by one set of bilingual workers, back translation into the original language by another set of bilingual workers, and

negotiating differences. He noted the importance of an anthropological model providing a more rigorous, systematic and contextual approach to translations.

From an anthropological perspective, the problem of translation is that the referents of symbols (their meanings) are aspects of a culture, not objects outside language through which language obtains meaning (Kleinman, 1987).

In a review article Guillemain et al. (1993) identified 17 studies of cross-cultural adaptations of English language instruments to measure health-related quality of life. In eight studies the translation techniques varied from one literal translation to three translations performed independently by one or two translators, back translations in six studies, one with multiple translations and back translations and five studies using committees comprising two to 12 bilingual individuals to review translations and back translations.

Qualitative techniques

Consensus methods (Fink, Kosecoff, Chassin and Brook, 1984; Jones and Hunter, 1995) have been used to determine the extent to which experts or lay people agree about a given issue (consensus measurement) and to resolve disagreement (consensus development) (Jones and Hunter, 1995). Since the late 1960s consensus methods have been used to develop consumer and professional definitions of the roles and qualities of primary healthcare (Fink et al., 1984), to measure task delegation among differing nursing skill levels (Trevdi, 1982), to elicit team judgement in the selection of quality assurance topics (Williamson, 1978), to examine appropriate indications for prostatectomy (Hunter and McKee, 1994) and other clinical interventions (Scott and Black, 1991a), in education and training (Mackie, Priest, Ginzler and Black, 1992), and practice development (Justice and Jang, 1990), and to identify measures for clinical trials (Felson 1993; Galleger, Hares, Spencer, Bradshaw and Webb, 1993).

The purpose of the present study was to test the feasibility of a combined qualitative and quantitative approach: a nominal group/expert panel to translate and assess the extent of agreement (consensus measurement) on the appropriateness of the translation and to resolve disagreement (consensus development).

Composition of the group/panel

In previous research the membership of nominal

groups has been selected on the basis of previously expressed interest in the topic (Hunter et al., 1994), representation of a variety of backgrounds (teaching, non-teaching hospitals, urban and rural district) (Scott and Black, 1991 b) and expertise in the subject (Hunter et al., 1994). Although a random sample is not a prerequisite for qualitative research (Mays and Pope, 1995) the sample should be selected to ensure that no particular interest or opinion is likely to dominate (Jones and Hunter, 1995). Highly educated individuals may not be representative of the target population (Secheres, Fay, and Zaidi, 1972) but some participants should be aware of the objectives of the material to be translated and the concepts involved so as to offer a more reliable restitution of the intended methods (Del Greco, Walop, and Eastrige, 1987). The nominal group leader is also a participant and should have subject-matter expertise (Fink et al., 1984).

Methods

Participants

Our panel comprised a total of nine individuals enrolled on the basis of fluency in both English and Sinhala. We preferred panellists to have some medical background and attempted to include doctors, but due to pressure of work this was not possible. The pilot was carried out in London using Sri Lankan professionals, including five doctors in a panel of 11, to translate the first three questions of the same questionnaire. This did not show any significant difference between the contribution of the doctors and that of the other subjects in terms of language expertise. We therefore recruited a group of students, including six medical students, who responded to an advertisement. On the grounds of expertise, the first author participated as one of the panellists and facilitated the group (Fink et al., 1984; Jones and Hunter, 1995).

Source instrument for translation

Each participant was provided with the Bradford Somatic Inventory (Mumford, 1992), for translation in advance of meeting. The Bradford Somatic Inventory is described as a multiethnic inventory of somatic symptoms (Mumford, 1991). The items were derived from the psychiatric case notes of Pakistani and indigenous British patients with clinical diagnoses of anxiety, depression, hypochondriasis or somatoform disorders ('neurotic spectrum'). A pilot version has been tested

in India and validated in medical clinics in Britain and Pakistan (Mumford, 1991). It has been used subsequently in the detection of psychiatric disorders among Asian patients presenting with somatic symptoms (Mumford, 1992).

This instrument was used in a controlled clinical trial of cognitive behavioural therapy for patients with medically unexplained multiple complaints, conducted in Sri Lanka by the first author. The justification for using this particular instrument and the results of that study is reported elsewhere (Sumathipala, Hewege, Hanwellla and Mann, in press).

Procedure

See Figure 1. Each participant was provided with the Bradford Somatic Inventory (Mumford, 1992) for translation in advance of meeting.

Translations were done individually. The panel then assembled for the next stage.

- The principles of this method (Fink et al., 1984; Scott and Black, 1991b; Hunter et al., 1994; Jones and Hunter, 1995) were presented by the first author and a discussion followed to clarify the process.
- Participants recorded their own translations of each question on separate flip charts. This was done in order to rate each translation individually without any comparison with others. Initials of the participant were noted on the back of the flip chart. Only one item of the original English version of the questionnaire and its corresponding nine translations were taken up at a time for evaluation and consensus generation.
- During the first round similar translations were grouped together.
- Each participant privately rated the appropriateness of the translation on a scale of 1 to 9 on the rating sheet prepared for the first round (Fink et al., 1984; Scott and Black, 1991b; Jones and Hunter, 1995). See Table 1 for details. Participants were asked to decide whether each translation represented the idea conveyed by the original item in English. We were concerned about the total meaning conveyed by the sentence rather than a direct word-to-word translation so linguistic equivalence was less important than conceptual and semantic equivalence. No discussion took place during this round. Each translation was rated individually without discussion. Rating sheets were collected at the end.

- The ratings were tabulated and presented. Panellists were presented with a summary of first round ratings. These included the median and range (dispersion) but not individual ratings (Hunter et al., 1994).
- Translations achieving consensus as appropriate, based on the predefined guidelines (Table 1) were submitted to the third and final round (Fink et al., 1984; Scott and Black, 1991; Jones and Hunter, 1995). Translations achieving consensus as inappropriate were excluded from further consideration.
- Translations failing to achieve consensus were submitted for discussion in the second round when amendments were made. Translations were regrouped if necessary and new ratings were made. Participants were also given a choice to rerate the translations during the second round, even if amendments were not made to the translation. It was made clear that the participants need not conform to the group view (Jones and Hunter, 1995).
- The second-round ratings were then tabulated and presented to assess the agreements/disagreements. Translations achieving consensus as appropriate were submitted to the third and final round to select the one with the highest preferences. Translations achieving consensus as inappropriate were excluded from further consideration. Translations failing to achieve consensus were also excluded from further consideration.
- Translations agreed as appropriate during the first and the second rounds were then relisted, discussed, and ranked to find the most satisfactory translation based on the consensus of the group. Ranking was done by individual confidential voting on a ranking sheet. The level of consensus required in the third round voting was decided in advance (Fink et al., 1984) as more than 50% support from the panel.
- All the above steps were repeated with each stem of the questionnaire.

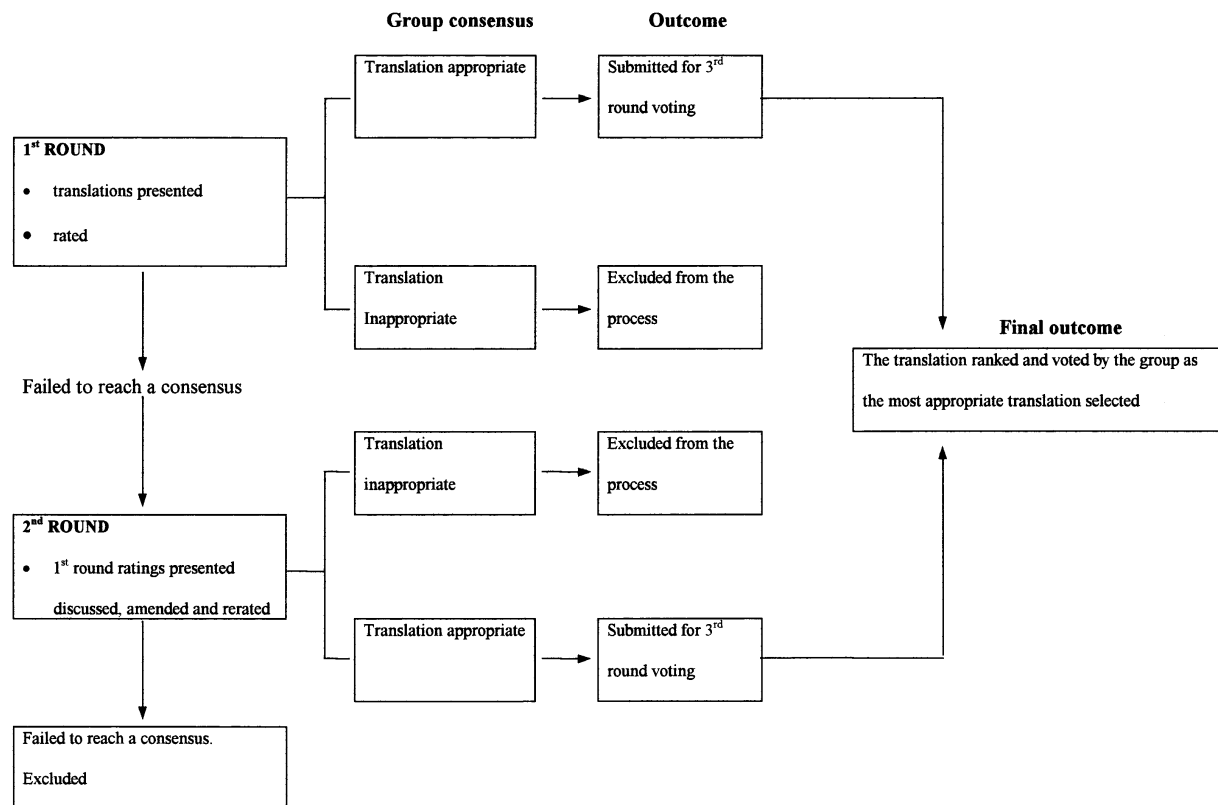


Figure 1: The process of consensus generation.

Table 1: Guidelines for rating and definitions on agreement

- If fully agree on the meaning conveyed by translation then rate 9.
- If totally disagree on the meaning conveyed by translation then rate 1.
- If the translation carries the same meaning or fairly satisfactory but requires changing some words, then rate between 5–8 depending on the appropriateness otherwise
- If you feel that the translation does not carry the same meaning or is not that satisfactory but some of the words used in the translation are useful, then rate between 2–5.

Agreement (consensus) (Hunter et al., 1994; Jones and Hunter, 1995)

Rating 1–3 – agree that the translation is inappropriate.

Rating 7–9 – agree that the translation is appropriate.

Equivocal

Rating 4–6 (Hunter et al., 1994).

Disagreement (Scott and Black, 1991)

After discarding the highest and lowest ratings or eliminating two ratings furthest from the median

- if at least one rating fell in 1–3 range when the other's range was at the other extreme
- if at least one rating fell in 7–9 range when the others range was at the other extreme.

In our work, both equivocal agreement and disagreement were considered as failing to reach a consensus.

Analysis of ratings

See Table 1. Appropriateness of the translation and consensus were assessed taking into consideration the median of the ratings and overall measure of dispersion (Scott and Black, 1991b). Outliers were eliminated by the method recommended by Scott and Black (1991b). Two ratings that were furthest from the median were eliminated. In the event of two rates being equidistant from the median, the rating that increased the degree of consensus most was retained.

Panellists were defined as being in agreement if all ratings fell within three points at either extreme (1–3 or 7–9) after excluding two outlying ratings (Scott and Black, 1991b).

Results

There was wide variation between items both in the translation and in the process by which consensus was achieved (see Figure 2). Only one item was selected in the first round to be included in the translated version having been translated almost identically by all participants.

On five items no consensus was reached on any of the translations. After discussion the panel managed to jointly produce one translation each for those five items that was acceptable to all participants. Where it was difficult to find a single word to denote a symptom we had to explain it more in the item.

There were seven items where one translation was selected for each in the 3rd round of voting. There were eight items for which translations selected for the third round were similar or identical so that they were combined. Consensus was 100% for each of the nine items finally synthesized.

Therefore out of 21 items finally selected for the translated version, only five items were contributed by individual panellists. All the others were contributed by more than one panellist. No single individual produced a significantly a greater number of selected final translations. The range of contributions was five to eight translations with median of seven. Three panellists contributed seven items, three panellists eight items, two panellists six items, and one panellist five items to the finally selected questionnaire items. The facilitator (AS) made no greater contribution than the other participants.

Although the level of consensus needed in the third round voting was decided in advance as more than 50% support from the panel for the selection of the item, for most of the items selected questionnaire, the support was 90–100%. Only one item received just over 50% (5/9) support.

Discussion

The process of translation in this paper confirmed the problems highlighted in the introduction of the linguistic evaluation protocol devised by the WHO (1997) for the ICDH-2 Beta field trial:

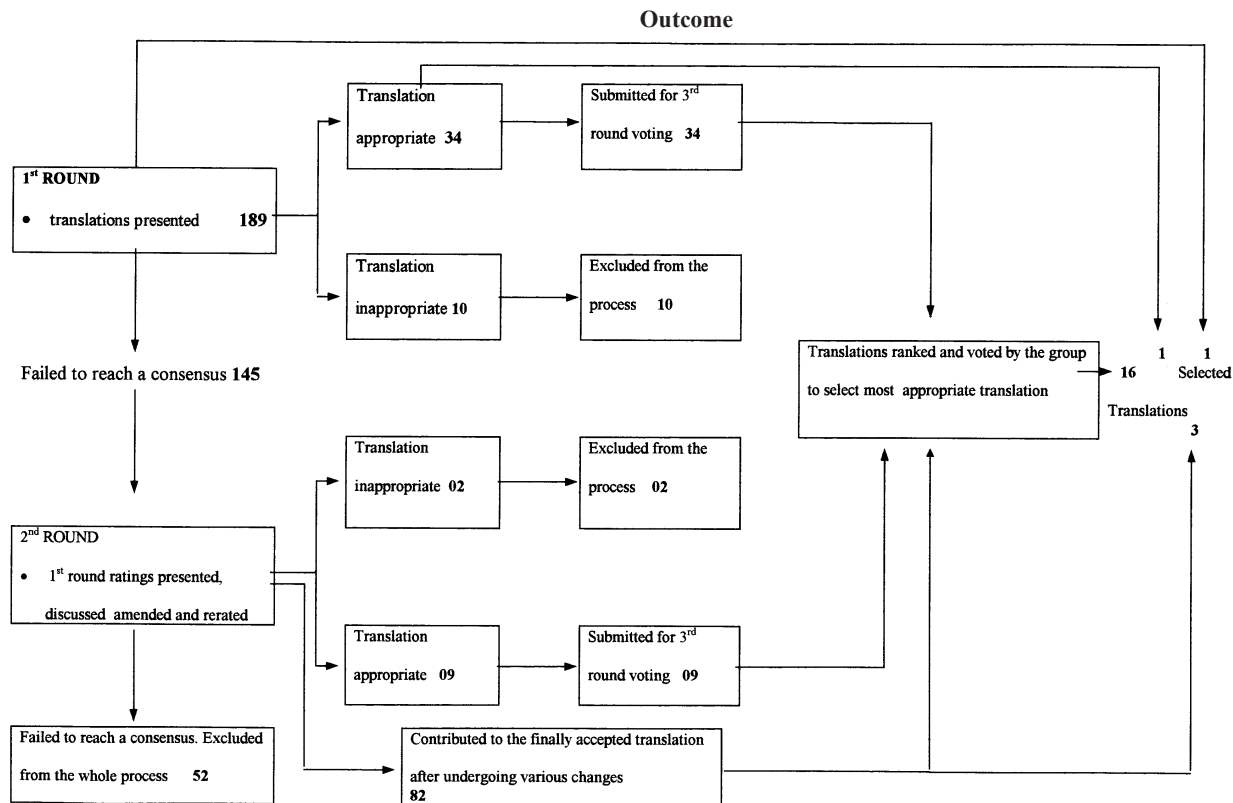


Figure 2: Outcome of the process.

- terms not being able to be translated into the local language; or
- the translation being very difficult because there is no exactly equal idiom or concept in the local language;
- having to modify the meaning of the original term during translation because only a part of the original meaning is present in the local term;
- losing some of the connotations, making the item too narrow; or
- having to expand the original meaning in the local language term as the local term has more (and different) connotations than the original, making the translated item too broad in meaning.

All the above findings suggest that the likelihood of one individual producing translations with a 100% probability of acceptance to a wider group is unrealistic.

Our study also confirms the opinion of Gullium et al. (1993) that quality will be higher if each translation is undertaken by teams rather than individuals, who are likely to introduce personal idiosyncrasies. The group managed to modify or eliminate inadequate or ambiguous items and to generate culturally appropriate semantic and conceptual equivalents. This is more important in the development of a research instrument than idiomatic equivalence.

It might be argued that the issues faced during the translation results from idiosyncrasies in the source instrument, which was developed in a different language and then translated to English, therefore making it more difficult to translate into another language. If this were the case, it would strengthen our argument for using a group to translate an instrument from one language to another.

This same argument can be applied to back transla-

tions by a different set of individuals. The issues faced during the forward translation are likely to occur in back translations, particularly the variation in translations by different individuals. Therefore, back translation does not necessarily guarantee English translations identical to the original English question and does not provide an adequate measure of appropriateness in a question from English translation from one language to another language. Although these differences could be negotiated by a few individuals or in a committee, the process would be less systematic with no measure of consensus.

Reliance upon back translation as a means of comparing the appropriateness of translation, will not produce a questionnaire with content validity for a particular culture. This problem was clearly demonstrated in the 'Spanish translation and cultural adaptation of the Diagnostic Interview Schedule for children in Puerto Rico' (Bravo et al., 1993). In the phobia section, an item concerned with travelling in trains did not have content validity in a country with no passenger trains. Similarly, with the item in our translation dealing with palpitation, back translation of the Sinhala version had no content validity in English cultures and vice versa. The appendix to this paper clarifies further how the content and semantic validity interfere with the translation process. 'Sinking' was not considered to have a semantic equivalent in the local language that could also retain the content validity of the translated item. Therefore the contents of the translated sentence had to be modified. On the other hand although 'wind' had a semantically meaningful equivalent in local language, that word was not considered to have validity in terms of its content. In summary, the process of translation confirmed that, at times, only part of the original meaning could be expressed in local terms. This resulted in the loss of some connotations making the item too narrow, or conversely in having to expand the original meaning in the local language terms making the translated item broader. Hence it is clear that in the translation process, establishing content validity would not necessarily guarantee a semantic validity or visa versa.

In the present study back translation was abandoned for two reasons. Firstly because we encouraged 'decentring' in order to promote conceptual equivalence (Drennan et al., 1991) and therefore it was not realistic to expect a back translation that is directly

comparable. Secondly diversity of translations by panel members occurred in the forward translation. If back translation is to be undertaken it should follow the same process of consensus generation as in forward translation.

Although the level of consensus needed in the third-round voting was decided in advance as more than 50% support from the panel for the selection of the item, in most of the items selected questionnaire, the support was 90–100%. Only item 19 received just over 50%(5/9) support.

Time factor

Important factors determining the feasibility of a particular technique are the constraints of time and resources. The preservation of an instrument is time consuming but unavoidable (Bravo et al, 1991).

Our questionnaire with 21 stems took four sessions to complete. Therefore the average time for each question was around one hour. This is cost effective.

Conclusions and recommendations

We have described the process of developing a method to assess and guide translation using a Combined qualitative and quantitative approach. This method not only allows for translation of instruments in a systematic and collective manner, reaching consensus on the appropriateness of the translation and subsequent amendments. This process is superior to arbitrary decision making by one or a few individuals handling the translations, because it quantifies the extent of agreement on the appropriateness of the translations.

When a different language is involved, translation and back translation are usually not sufficient to obtain cultural equivalence with the instrument (Cainno et al., 1997).

Pope and Mays (1995) have recommended using qualitative techniques to discover the most comprehensible terms or words to use in a subsequent questionnaire. Our findings confirm the feasibility of extending the principles of one qualitative method, the nominal group or expert panel, for translation of whole sentences in questionnaires.

Although we have translated and adopted a questionnaire, we recommend experimenting with this method for adapting not only questionnaires but also other instruments, including structured or semi-structured interviews, because adaptation of an instrument

to a different culture is some thing beyond a mere translation attempt. The method could be strengthened by convening another panel with different participants to discuss the items that caused most difficulty and in reaching a consensus. Researchers may also use more than one panel to select translations for the final round. Questionnaires could be translated using separate panels with different categories of people having different viewpoints (doctors, lay people and patients) and combine the translations selected for the final round in order to generate consensus. The percentage of support required in the final round to select items can be set at higher than 50% to be more selective in the choice.

On the basis of this study, we recommend further use of this method as a more rigorous, systematic and contextual approach to the translation of research instruments.

Appendix

Some examples of difficulties encountered in the translation process

Q 12. None of the translations managed to gain consensus during the first round but almost all of them were very similar. During the second-round discussion it transpired that failing to reach a consensus mainly arose from the translation of the word 'heart pounding'. The group came out with two words (*dada bada* and *dug dug*) to denote pounding. However these had absolutely no meaning in back translation. In colloquial language *dada bada* indicate speeding and *dug dug* indicate a feeling of a forceful tapping and incorporating a form of a noise with in it. Both these words are used by laymen to describe increase heart rate. The word *dug dug* was preferred. After rephrasing with this addition it was possible to group them together and all agreed on the final synthesis.

Q 19. 'Has your heart felt weak or sinking?' All nine translations failed to gain consensus. The panel had greatest difficulty in finding a culturally appropriate, conceptually meaningful word for 'sinking'. Direct translation with a linguistic equivalent of this word was not appropriate. Ultimately the group agreed to ignore 'sinking' and decided only to translate the remaining part of the sentence.

Even this translation, the group did not feel having a culturally meaningful conceptual validity. However inclusion of this will enable us to see whether the patients will make any response to this item.

Q 20. 'Have you suffered excessive wind or belching?' None of the nine translations managed to establish consensus, but after discussion the panel managed to jointly produce one translation that was acceptable to all. Both words 'wind' and 'belching' created problems in this translation. 'Wind troubles' are a common attribution of symptoms by Sri Lankan and Asian patients. Therefore the direct translation of the question would have made an impression on the patient that the question is about 'suffering due to wind trouble'. But in the original English version it meant wind in the gut, *flatus/farte*. It was difficult to find a single word to denote it. Instead we had to explain it. *Flatus/farte* has a single colloquial word understood by ordinary people, but that word was not considered by the group as civilized enough to use in a questionnaire. The word 'belching' has a dictionary translation but that is not a familiar term to everybody. This called it *ramathel*. *Thel* is oil but *ramathel* has nothing to do with oil. We therefore had to describe it and with brackets inserted around *ramathel*.

Acknowledgements

Wellcome Trust, UK, for funding the main research project for which translation of the instruments was needed. Professor Anthony Mann, Section of Epidemiology and General Practice, Institute of Psychiatry for his supervision and support for the whole project and for his comments on this paper. We wish to thank members of the nominal group, Dumindu Witharana, M.S. Oowise, Indu Karangoda, R. Kithsiri, Sisith Kulesekera, H. Ariyapperuma, Nuwan Samarakkody, and Suvin Hewege (non-clinical research assistant) for their valuable contribution.

References

- Bravo M, Woodbuyr-Farina, M, Canino, J, Rubio-Stipec M. A cross cultural adaptation of a psychiatric epidemiological instrument. *Cul. Med. Psychiatry* 1991; 15: 1-18.
- Bravo M, Woodbuyr-Farina M, Canino J, Rubio-Stipec M. The Spanish translation and cultural adaptation of the diagnostic interview schedule in Puerto Rico. *Cul Med Psychiatry* 1993; 17: 329-44.
- Brislin, RW, Lonner W, Thorndike R. *Cross Cultural Research Methods*. New York: John Wiley & Sons, 1973.
- Cannon G, Fernandez RL, Bravo M. Methodological challenges in cross-cultural research. *Transcultural Psychiatry* 1997; 34(2): 163-73.
- Cannino G, Bird H, Rubo-Stipe M, Bravo M. The epidemiology of mental disorders in adult population in Puerto Rico. *P R Health Sci J* 1997; 16(2): 117-24.

- Del Greco L, Walop W, Eastrige. Questionnaire development. *Can Med Assoc J* 1987; 136: 817–18.
- Drennan G, Levett A, Swarts L. Hidden dimension of power and resistance in the translation process: A South African study. *Cul Med Psychiatry* 1991; 15(3): 361–81.
- Felson DT. Choosing a core set of disease activity measures for rheumatoid arthritis clinical trials. *J Rheumatol* 1993; 20: 531–4.
- Fink A, Kosecoff J, Chassin M, et al. Consensus methods: characteristics and guideline for use. *Am J Public Health* 1984; 74: 979–83.
- Galleger M, Hares T, Spencer J, Bradshaw C, Webb I. The nominal group technique: a research tool for general practice. *Fam Pract* 1993; 10: 76–81.
- Gaviria M, Pathak DS, Michell T, Flaherty JA, Winthrop RM, Martinez H, Pacheco GC, Richman J, Birz S. Designing and adapting instruments for a cross cultural study on immigration and mental health in Peru. American Psychiatric Association, 1984.
- Guilliman F, Bombardier C, Beaton D. Cross cultural adaptation of health related quality of life measures: literature review of proposed guidelines. *J Clin Epidemiol* 1993; 46(12): 1417–32.
- Hunter DJW, McKie CM, Sanderson CFB, Black NA. Appropriate indications for prostetectomy in the UK. Results of a consensus panel. *J Epidemiol Community Health* 1994; 48: 58–64.
- Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995; 311: 376–80.
- Justice J, Jang R. Tapping employs insight with the Nominal Group Technique. *Am Pham* 1990; 30: 43–5.
- Kleinman AA. Anthropology and Psychiatry. The role of culture in cross cultural research on illness. *Br J Psychiatry* 1987; 151: 447–54.
- Makee M, Priest P, Ginzler M, Black N. Which tasks performed by pre registration house officers out of hours are appropriate? *Med Edu* 1992; 26: 51–7.
- Mays N., Pope C. Rigour and qualitative research. *BMJ* 1995; 311: 109–12.
- Mumford DB, Bavington JT, Bhatnagar KS, Hussain Y, Mirza S, Naraghi MM. The Bradford Somatic Inventory: a multi-ethnic inventory of somatic symptoms reported by anxious and depressed patients in Britain and the Indo-Pakistan subcontinent. *Br J Psychiatry* 1991; 158: 379–86.
- Mumford DB. Detection of psychiatric disorders among Asian patients presenting with somatic symptoms. *Br J Hosp Med* 1992; 47: 202–4.
- Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *BMJ* 1995; 311: 42–5.
- Scott EA, Black NA. Appropriateness of cholestectomy – a consensus panel approach. *Gut* 1991a; 9: 1066–70.
- Scott, EA, Black NA. When does consensus exist in expert panels? *J Public Health Med* 1991b; 13: 35–9.
- Secheres L, Fay TL, Zaidi SMH. Problems of translations in cross cultural research. *J Cross-cultural psychology* 1972; 3: 41–56.
- Sumathipala A, Hewege S, Hanwella DRC, Mann AH. Randomised controlled trial of cognitive behavioural therapy for repeated consultations for medically unexplained complaints: a Sri Lankan study. *Psychol Med* (in press).
- Trevdi VM. Measurement of task delegation among nurses by nominal group analysis. *Med Care* 1982; 20: 154–64.
- Werner O, Campbell DT. Translating, working through interpreters, and the problem of decentering. In: *A Handbook of Methods in Cultural Anthropology*. New York: American Museum of Natural History Press, 1970.
- Williamson JW. Formulating priorities for quality assurance activity. *JAMA* 1978; 239: 631–7.
- World Health Organisation. International Classification of Impairments, Disabilities and Handicap (ICIDH). Field trial protocol for linguistic analysis. Geneva: WHO, 1997.

Correspondence to Athula Sumathipala, Section of Epidemiology and General Practice, Institute of Psychiatry, Kings College, London. SE5 8AF. Fax 44-020-72770283. Email spjuats@iop.bpmf.ac.uk.