

So you want to do research?

5: Questionnaire design

Keith A Meadows

Collecting information using a questionnaire as part of a research study, service evaluation or audit, is now common practice. There is a variety of questionnaire types that can be used. Whichever way questionnaires are used and for whatever purpose, the same design rules should apply in their construction to ensure that they are appropriate for their intended use. Each questionnaire should be designed to minimize respondent and interviewer errors in the understanding of the questions and recording of the answers, as well as maintain the interest and cooperation of the respondent. Questionnaires are often used with little thought given to these issues, which can lead to the collection of unreliable information.

This paper, the fifth in the series, describes the different types of questionnaire together with their advantages and disadvantages, and the key issues of questionnaire design including content, question construction, questionnaire format and adapting existing questionnaires.

ABSTRACT

This article describes the key aspects in the design, construction and adaptation of survey questionnaires. There are different types of questionnaire, each of which has its advantages and disadvantages. Aspects of constructing the questionnaire are discussed in detail; choosing the mode of administration; the objectives of the survey; availability of resources; characteristics of the target population; and quality of data. Issues concerning the identification of the questionnaire's content, wording and sequencing of the questions, through to the overall appearance and layout of the questionnaire are also considered. Differences in the role of open-ended and closed questions, together with their strengths and weakness, are outlined, and the need to undertake pre-testing and piloting as an integral part of questionnaire development is highlighted. Finally, issues around the adaptation of existing questionnaires are discussed with particular emphasis on their use in different language and cultural groups, and the need to achieve conceptual, content, semantic, operational and functional equivalence is described. An overview of the translation process is provided.

Types of questionnaires: advantages and disadvantages

The first choice that needs to be made when designing a study or survey, is the mode in which the questionnaire will be administered.

Self-completion questionnaire

Self-completion questionnaires are completed in writing by the respondent. The most common use of self-completion questionnaires is delivery and return through the post. However, they can also be completed in the presence of the researcher (supervised self-completion), who can provide assistance and check the questionnaire for completeness.

Self-completion questionnaires offer low cost relative to other methods, potential coverage of a widely dispersed population and a wider coverage in the study population. They avoid the possibility of interviewer bias, although weaknesses in design and wording can still lead to biased reporting.

However, self-completion questionnaires have the potential for low response rates, require a level of literacy to complete the questionnaire as well as the availability of an accurate list or sampling frame from which to select the sample of respondents. Self-completion questionnaires are appropriate for less complex topics and need to be easy to complete without assistance. It is generally suggested that self-completion questionnaires be shorter than those administered during interview and contain mostly closed ended questions (Bourque and Fielder, 1995). They are less suitable for recording open-ended data, complex questioning and question skip patterns (in which the answer to one question dictates which subsequent questions are answered). Bourque and Fielder (1995) provide a more detailed discussion on the advantages and disadvantages of self-completion questionnaires.

Interviewer-administered questionnaire

With interviewer-administered questionnaires, each respondent is asked the same questions by the interviewer, in the same way, in order to eliminate as far as possible any bias. Advantages of this mode of administration include the collection of more

detailed and complex data, the possibility to clarify misunderstandings and the opportunity for the interviewer to probe for additional information. Open-ended questions can be used, in addition to filter questions (see above and *Box 4*) and complex question skip patterns. Recording of the information is undertaken by the interviewer, so is not dependent on the respondent's level of literacy.

Interviewer-administered questionnaires are generally more costly and resource-needy than self-completion questionnaires, because interviewers often need to be trained and their administration may involve travel. They can also be more time consuming. The personal characteristics of the interviewer – e.g. age, class, sex, race, level of experience – can have an effect on the response rates and the nature of the responses. Interviewers can also introduce both random and systematic error, e.g. in the recording of answers, by changing the wording of the questions or by selective recording of answers and differential probing between interviewers (Oppenheim, 1992).

Telephone interview

In some ways, telephone interviews maximize the advantages of self-completion questionnaires and face-to-face interviews, while minimizing their disadvantages. Because the need to travel is removed, they can cover a widely dispersed population and achieve a wide coverage in the study population relatively quickly and at a low cost. Response rates, although generally not as high as face-to-face questionnaires, can be between 90% and 95% (although they are dependent on the topic of the survey, with higher rates when the topic is of direct interest to the respondent) (Morton-Williams, 1993). Interviewer characteristics such as age and race will have a less direct effect on the answers given than other variables, e.g. accent. Telephone interviews are considered suitable for most complex subjects, are less likely than face-to-face interviews to obtain just socially acceptable answers, and can deal with more sensitive subjects.

An obvious disadvantage of telephone interviews is the problem of sampling bias and generalizability of the findings, as people of lower income, young people, and ethnic minorities are less likely to have a telephone. They are also unsuitable for samples which comprise people who are hard of hearing, older people and some ethnic minorities, unless the interviewer is speaking the same language. Visual aids such as prompt cards cannot be used and the use of complex response formats and questions with long lists of multiple choice answers should be avoided because of memory effects. Detailed discussion on interviewing and telephone surveys can be found in Frey and Oishi (1995) and Oppenheim (1992).

Constructing a questionnaire

Constructing a questionnaire and wording the questions are not simple tasks. They require skill and an understanding of the key issues, as well as time to develop a questionnaire that is appropriate to meeting the objectives of the study.

One of the major pitfalls that the novice, and sometimes expert, falls into when designing a questionnaire is formulating questions that are difficult to ask or answer. These may include those that use unfamiliar words or phrases, or are too vague and unspecific, e.g. 'How have you been feeling today?'. Double-barrelled and catch-all questions are also unsatisfactory, e.g. 'How satisfied are you with the medication you are taking and the staff who look after you?'. Overly long questions are difficult to answer, e.g. 'Over the past year have you attended any hospital including your local hospital, but not including accident and emergency departments or to visit friends or family?'. Questions that invite distortion or are leading should also not be used, e.g. 'Do you think patients should be examined by a doctor of the same sex?'.

Reliability and validity

The practical value of a questionnaire depends on the reliability and validity of the information it collects.

Reliability refers to how well data collected by using a questionnaire can be reproduced. The most common indicator of reliability is test-retest reliability. This is a measure of how stable the respondents' responses are between time 1 and time 2, when we can assume that there should be no natural change in the responses, e.g. because of treatment, maturation etc. A poorly-designed questionnaire might result in variation in the responses of the respondents between times 1 and 2, leading to measurement error. The common method of measuring test-retest reliability is a statistical test – the correlation coefficient between the two sets of responses, which should not be less than 0.70 (Litwin, 1998). Other tests of reliability include measuring the interviewer's consistency and the consistency between interviewers.

Internal consistency is another measure of reliability but is different from those mentioned above; it is a measure of how well a group of questions 'tap' a particular concept. For example, we would expect a scale designed to measure physical ability to comprise items that tap the various aspects of the concept, e.g. the ability to climb stairs, walk a block, run 100 metres etc. If this were the case, the scale would then have a high internal consistency.

Validity is how well the questionnaire measures what it is intended to measure. For a measure to be valid it must be reliable, but something can be reliable without being valid – a clock that is always

'The practical value of a questionnaire depends on the reliability and validity of the information it collects.'

‘Findings from high-grade studies indicate that no particular mode of administration is superior in all respects or in all settings, and that the choice of mode should be decided on a survey-by-survey basis.’

10 minutes fast is reliable, but is not a valid indicator of time. Validity of a questionnaire can be measured in a number of ways, including face, content, criterion and construct validity.

Face validity is the least scientific and is based on a basic review of whether the items look appropriate. If we were asking about attitudes to smoking then the measure would have face validity if the questions appeared to be about attitudes to smoking.

Content validity, which is often and incorrectly confused with face validity, is assessed on the extent to which the questionnaire’s content includes everything it should, and does not include anything it should not. Evaluating content validity of a questionnaire should be based on expert review, e.g. expert panels including patients.

Criterion validity is how well the questionnaire is able to predict some future event, behaviour or outcome, or how it compares with a similar measure of the same thing.

Construct validity is the most difficult form to assess. It is often based on the extensive use of the questionnaire and is the amalgamation of all the evidence of its performance, including content and criterion validity.

Wherever possible it is recommended that existing questionnaires that have been widely used and been shown to be reliable and valid, should be used. This is often possible when we wish to use standardized questionnaires that have undergone considerable development and about which there is published evidence of their reliability and validity. However, this is not always possible, particularly when the information requirements of the questionnaire are specific to a one-off study. Under such circumstances the only option is to either adapt an existing questionnaire or construct a questionnaire from scratch, which can include questions adopted from existing questionnaires wherever possible.

A detailed discussion on all the key aspects of questionnaire design and construction is beyond the scope of this article, but a number of excellent books and other publications are available which provide an in-depth discussion of the topic. These include, Dillman (1978); Sudman and Bradburn (1982); Oppenheim (1992); Fowler (1995); Litwin (1995); and Jenkins and Dillman (1997).

Choosing the mode of administration

Deciding on the mode of administration is the first stage in choosing or developing a questionnaire. The advantages and disadvantages of the different modes of administration need to be taken into account. Findings from high-grade studies indicate that no particular mode of administration is superior in all

respects or in all settings, and that the choice of mode should be decided on a survey-by-survey basis (McCull et al, 2001). The factors which should be taken into account include the following:

- If the *objectives* of the survey are, for example, to gain an insight into the attitudes of homeless people, then a self-completion or telephone-administered questionnaire is not feasible. If, however, we wanted to find out the level of satisfaction of a sample of patients registered with a general practice using a five-item questionnaire, this could be carried out using either a self-completion or a telephone interview.
- Availability of *resources* include costs, personnel to conduct interviews or handle administrative tasks, skill and time to develop and administer the questionnaire, data coding and entry, data analysis.
- Characteristics of the *target population* include their overall educational level and the accessibility to the respondents. If the potential respondents are geographically widely dispersed then face-to-face interviewing will be impractical. If the target population is difficult to access, e.g. homeless people, or is considered not to have an overall educational level sufficient to understand and answer questions spoken over the telephone, then self-administration or telephone interviewing is likely to be inappropriate and logistically difficult.
- The *mode of administration* can have an impact on the quality of the data generated. This includes sampling method and the ability to reach all eligible respondents, response rates, confidentiality, interviewer effects, the volume and complexity of the data collected.

Questionnaire content

Tailoring the content to meet the objectives of the survey ideally includes both a comprehensive review of the literature and the undertaking of a small-scale qualitative study to identify the range of behaviours, attitudes and issues relevant to the objectives of the survey. A literature review will provide information on any other studies which have been done on the topic, the current state of knowledge in the area, help build on or extend current work, identify methods of data collection, and highlight the content of existing questionnaires and any problems experienced (Bourque and Fielder, 1995). Qualitative research enables the exploration of views and behaviour patterns, using the two main methods of in-depth interviews and focus groups. It can provide the basis for deriving relevant questionnaire content (Meadows, 2003).

The topic areas and variables, and their relationships with other variables identified from the litera-

ture review and qualitative research, need to be included in the questionnaire. They should be listed and decisions should be made on how they can be measured. From this, the detailed design work can start, such as formulating the specific questions, deciding on the sequence of questioning and listing answer categories, adding instructions, and signposts and question skip instructions if its an interview administered questionnaire (Hoinville and Jowell, 1978). One approach to achieve this is to develop a flow chart of the different questions and sections of the questionnaire (Figure 1).

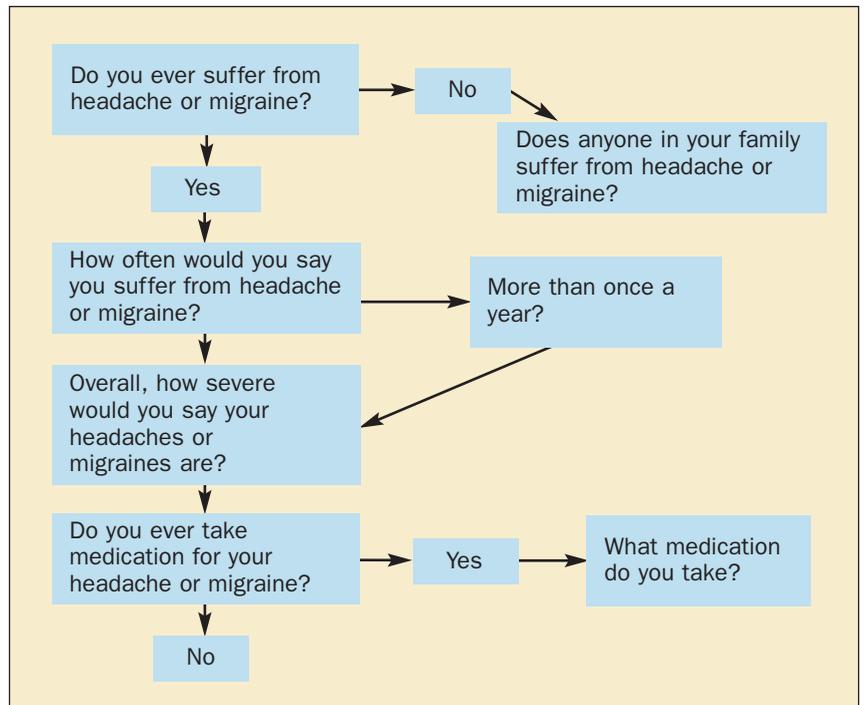
Open/closed questions

Open-ended and closed questions both have their place in survey design. A closed question (Box 1) provides a choice of alternative answers from which the respondent is asked to select, by ticking or underline, or choosing from a read-out list or prompt card. Closed questions can be attitudinal or factual and the choice of answers or response options form very much part of the question (Oppenheim, 1992).

Closed questions enable comparison across individuals or groups of respondents, require less time to complete than open-ended questions and are easy to code and process. However, they cannot capture in-depth or spontaneous responses and can sometimes unknowingly bias answers by forcing the respondent to choose between alternative responses or consider options that they had not previously considered (Oppenheim, 1992).

In contrast, open-ended questions do not provide any predetermined answers, which enables the respondent to answer the question using his or her own words. This can be very useful when trying to identify in more depth the respondent's thoughts, feelings and experiences. Open-ended questions provide opportunities to probe and can also help to identify the range of answers or response categories for use with a closed question. As a consequence open-ended questions are important in the development stage of a questionnaire. One disadvantage is that they require more effort from the respondent and therefore should be used sparingly in self-administration questionnaires. Furthermore they are subject to interviewer variability, are more time consuming to answer, and consequently limit the number of questions that can be asked in a given time. Answers to open-ended questions are more difficult to analyse than those to closed questions and make comparisons between groups difficult.

Which is more valid? An open-ended question can produce more non-common category responses than a closed question, but these tend to be small and miscellaneous and providing the response categories



in the closed question cover the main issues identified in previous developmental work, the use of either question will result in similar conclusions (Schuman et al, 1986; McColl et al, 2001).

Figure 1. Flow chart for questioning sequence.

Question wording and sequencing

Miller (1984) has described the wording and ordering of questions as the 'rich folklore of survey research'. Constructing a questionnaire must be seen as an important and integral part of the whole research process.

The principle aim in writing a question is to ensure that each question means the same to the surveyor and respondent, who should be able to respond with as accurate a response as possible (Frey, 1989). Sudman and Bradburn (1982) have identified a number of factors that need to be addressed when designing survey questions. These include:

- **Memory.** Avoid over-taxing the respondent's memory
- **Motivation.** Ask questions that are relevant to the respondent
- **Communication.** Ensuring what we are asking

Box 1. Example of a closed question

Which of the following would you find most difficult doing? (Please circle the appropriate number)

- | | |
|---------------------------------------|---|
| Walk across the room | 1 |
| Climb one flight of stairs | 2 |
| Walk one kilometre | 3 |
| Run half a kilometre without stopping | 4 |

‘Common sense should tell us that paying attention to the design and layout of a questionnaire is an important stage in its development.’

the respondent is clearly communicated

- **Knowledge.** Only ask for information the respondent is likely to have.

Despite the wide range of research and the evidence base for ‘best practice’ in questionnaire design, relatively little of this can be generalized to health-related research and so caution should always be exercised when extrapolating findings to health surveys (McColl et al, 2001). Nevertheless, there are recognized guidelines and principles of question wording which should be followed whatever the chosen mode of administration (*Box 2*).

Response categories

The choice, wording and ordering of response categories is as critical as the wording of the question itself and can have a significant impact on how respondents interpret and answer the question. As with evidence on best practice for question wording, findings are inconsistent. However, there are a number of common-sense recommendations, as well as recommendations for practice based on one or more high grade comparative studies (McColl et al, 2001):

- Ensure response categories for closed questions are mutually exclusive, i.e. do not overlap (e.g. What is your age next birthday? 15–25, 26–35, 36–45 years, as opposed to 15–25, 25–35, 35–45)
- All potential options/categories are exhaustive and if necessary include the option ‘other, please specify’
- Be aware that response options can send a message about the type and range of ideas, in addition to concepts the respondent should be thinking about

Box 2. Aspects of good question wording

- Use simple language, avoid the use of technical terms, professional jargon and abbreviations
- Avoid words that may have more than one meaning (e.g. dinner)
- Avoid questions that are not sufficiently specific (e.g. How well do you manage your diabetes?)
- Avoid ambiguity (e.g. What kind of day has it been for you?)
- Keep the maximum number of words in a question to approximately 20
- Avoid double-barrelled questions (e.g. How do you feel about the treatment you have been receiving for your present illness, and the doctors who are treating you?)
- Avoid double negatives – a negative statement followed by a disagree response (e.g. I feel there is no one I’m close to – disagree)
- Avoid leading questions (e.g. Many patients now days think that waiting times in the NHS are too long. Do you?)
- Avoid the use of loaded words and concepts (e.g. those which are emotionally coloured and suggest a feeling of approval or disapproval)
- Avoid the use of presuming questions (e.g. How long did you have to wait the last time you visited your local accident and emergency department?)

After Moser and Karlton (1971), Oppenheim (1992), (McColl et al (2001)

- An open space should be provided in self-completion questionnaires for free comment, which can improve response rates
- For factual questions the ‘don’t know’ option can be omitted.

Question sequencing

The position of a question may affect the way the respondent answers it (Schuman et al, 1983; Serdula et al, 1995). Researchers need to be aware of the potential effects of the order of questions in self-completion and interviewer-administered questionnaires. Researchers should follow recommendations on questionnaire design, including:

- Specific questions should follow on from general questions
- Demographic questions (e.g. age, sex, education and race) should be placed at the end of the questionnaire
- The apparent relevance and salience of opening questions can influence respondents’ motivation to complete the questionnaire. Therefore, relevant and salient questions should be placed close to the front of the questionnaire although highly sensitive questions should not be among the first few
- Whenever possible questions should be ordered from easy to difficult in the questionnaire
- Questions should be blocked by topic (McColl et al, 2001).

Formatting the questionnaire

Common sense should tell us that paying attention to the design and layout of a questionnaire is an important stage in its development. A well-designed questionnaire can simplify the tasks of the respondent, interviewer and data processor (Sudman and Bradburn, 1982), reduce errors in the asking of the questions and minimize variability between interviewers and data processors (McColl et al, 2001). Appearance and layout of a questionnaire can influence a respondent’s decision on whether to respond.

The main issues around the appearance and layout of questionnaires include length of questionnaire, question and response category format, print details such as font and typeface, pagination and instructions. However, unlike issues of question wording and sequencing, less attention has been paid to principles for the formatting of questions, or comparative studies on influence of questionnaire appearance and layout on response rates and bias. Nevertheless, expert opinion recognizes the importance of enhancing the appearance and layout of questionnaires, and is supported by psychological theories in a number of cases (McColl et al, 2001).

Length

Findings in relation to the effect of questionnaire length on response rates are equivocal, but a longer questionnaire may potentially lead to fatigue or carelessness, or put off potential respondents. However, reviews have shown that questionnaires on topics which are relevant or interesting can probably be longer than questionnaires on more general topics or those for the general population (Oppenheim, 1992; McColl et al 2001).

Question and response category format

Questions and response categories can be formatted horizontally or vertically (Box 3). Bourque and Fielder (1995) prefer the vertical format because the question is clearly differentiated from the responses, and also because it can make data entry simpler and less prone to error. Dillman (1978) and Sudman and Bradburn (1982) recommend the vertical format because it gives a less cluttered appearance and adds to the respondent's feeling of accomplishment. However, on the grounds of conserving space, an exception to the rule would be where a set of questions uses the same set of responses. Whether the respondent is asked to tick a box or circle a number appears to not be significant (McColl et al, 2001), but a mixture of circling and ticking should be avoided.

A consistent format should be kept throughout the questionnaire and the natural reading style of left to right should be considered when placing headings, codes or instructions. Whenever possible graphics should be used to steer the respondent or interviewer through the questionnaire or indicate skip patterns (Box 4).

Print details

Font sizes less than 10 points should be avoided. If it is anticipated that respondents might have visual problems, e.g. older people or particular illness groups, then 14 to 16 points should be considered, depending on the typeface. Over-use of upper case letters and italics should be avoided.

Pagination

There is little evidence as to whether questionnaires in the form of booklet or individual pages has any effects on response rates, or what the relative benefits are of double vs single-sided printing. However, McColl et al (2001) recommend using a booklet format with double-sided printing using a standard A4-folded-A5 format.

Instructions

Both self-administration and interviewer-administered questionnaires will almost certainly require

Box 3. Examples of a survey question formatted horizontally and vertically

Horizontal

How would you rate your health over the past 7 days?

Excellent	V. good	Good	Fair	Poor
1	2	3	4	5

Vertical

How would you rate your health over the past 7 days?

Excellent	1
V. good	2
Good	3
Fair	4
Poor	5

instructions. The purpose of instructions is to facilitate the process of collecting information by helping the respondent and interviewer move through the questionnaire. Self-administration questionnaires will generally include brief introductory comments on the purpose of the survey, the type of questions to be asked and why the information is required, in addition to instructions on what to do with the questionnaire when completed. These instructions should always be placed at the beginning of the questionnaire. However, more specific instructions relating to particular questions should be placed as close to the question as possible. Instructions for the interviewer are likely to include which questions should be asked of which respondents, the script asking the question and what probes should be used to get additional information from the respondent.

Other aspects of questionnaire appearance

- Avoid splitting a question or its response options across two pages.
- Do not use lines for responses to open-ended questions; leave sufficient white space between questions
- Consider the use of coloured paper to distinguish between different questionnaires
- Ensure the front cover contains the title of the survey and the name and address of the organization carrying it out
- Provide a back cover with blank space for respondents to make open comments
- Print a 'thank you' after the last question and provide details of where to return the questionnaire.

Box 4. Example of questions to guide the participants

ASK ALL: Q4. What is your current marital status?

Married	1	
Widowed	2	→ SKIP TO Q7
Single	3	
Separated or divorced	4	

'In the same way that health and illness may be conceptualized in different ways across cultures, the validity or relevance of the questionnaire items representing a given domain or concept may also vary.'

Pre-testing/piloting

Given the complexity of the questionnaire design process, it is highly unlikely that the first draft of a questionnaire will be perfect. Pre-testing the questionnaire can highlight any problems with it, including excessive length, incomprehensibility, missing questions etc. Pre-tests can be carried out using focus groups (Bourque and Fielder, 1995) and, more recently, cognitive aspects of survey methodology (CASM), which draws on the theories of cognitive psychology and the use of cognitive laboratory techniques to improve questionnaire design (Sirkin et al, 1999; McColl et al, 2003).

Pilot studies focus on testing the whole administrative procedure of using the questionnaire in a smaller but representative sample of the participants before the main study. Here the aim is to test the whole questionnaire, letter of introduction, instructions to participants, reminder letters etc. It is a small-scale test of the main study to check that all the procedures are working properly and, if not, to rectify them before the main study.

Adapting existing questionnaires

Rather than develop a new questionnaire, it is sometimes possible to use or adapt an existing one, or use some of its questions. While this may enable comparison with other studies, there are a number of considerations when adapting another questionnaire because the original is too long, or it is going to be used on a different population. First, adapting an existing questionnaire for a different purpose or group than that for which it has been developed can have serious implications for its reliability and validity. Because a questionnaire is reliable and valid in one setting, it cannot be assumed this is the case in all settings. Even a slight alteration to the wording of a question or the order of questions can impact on how people answer, so laying claim to the original questionnaire's reliability and validity should be avoided after adaptation. Before any adapted questionnaire is put into the field, it should undergo some pre-testing to evaluate its reliability and validity, and it is these findings which must be reported.

Some questionnaires are under copyright and the first step before either using or adapting the questionnaire is to contact the author to obtain permission to do so. Even if not under copyright, the user is obliged to notify, and respect the recommendations of, the author, and to cite the original source.

One reason why a questionnaire might require adaptation is where it needs to be used with people of different cultures and languages than for whom it was originally developed. When this is the case, there are

very strong arguments against the assumption that translating the content of the questionnaire into the required language will be sufficient (Herdman et al, 1998; Meadows and Wisher, 2000). Emphasis should be placed on achieving equivalence across the different cultural groups in the concepts measured and the items tapping these concepts. The different types of equivalence are outlined in *Figure 2*.

Conceptual equivalence

Investigating conceptual equivalence involves exploring the ways in which health and illness are conceptualized, as well as the values that are placed on them, and understanding that such meanings and values are part of the culture's social reality. An illustration of this is the work of Howlett et al (1992) on the analysis of the responses to the UK national health and lifestyle survey in an attempt to determine ethnic patterns in the concepts of health and illness. Comparing the beliefs of groups of Asians and Afro-Caribbeans with a matched sample of whites, they found that the Afro-Caribbeans were more likely to describe health in terms of strength and fitness, whereas Asians saw it in relation to one's ability to perform everyday activities.

How individuals experience an illness is a cultural phenomenon reflecting beliefs about aetiology, illness behaviour and the assigned roles of the respective parties (Hunt, 1986). Reluctance to express forms of emotional distress can be observed in a number of cultures for a variety of reasons.

Item/content equivalence

In the same way that health and illness may be conceptualized in different ways across cultures, the validity or relevance of the questionnaire items representing a given domain or concept may also vary. Item or content equivalence is established when the items estimate the same parameters or describe a phenomenon in each culture.

Item relevance will vary across cultures with respect to specific social and leisure activities or illness behaviour. For example, questions about difficulty in climbing stairs or ability to tend the garden will have little relevance in a culture where a large proportion of the population live in housing without stairs or a garden. Care should also be taken with items that could be considered as universal, such as the activity of dressing, but where inability to do so might be perceived as less or more serious (and indicative of differing levels of incapacity) in different cultures (Katz et al, 1963; Herdman et al, 1998). Questions will also vary in terms of acceptability, e.g. questions relating to suicidal ideas may be offensive to some cultures.

Semantic equivalence

Semantic equivalence is about retaining the meaning of each item after its translation into the target language(s), and is a key issue in achieving culturally equivalent questionnaires. Differences between languages and cultures in the salience of concepts, idioms and colloquialisms mean a literal translation from one language to another can be inappropriate. Literal translation, while allowing for changes in the word order and syntax, maintains a one-to-one correspondence between the words, possibly resulting in an incongruence in the meaning of concepts between the different languages (Bullinger et al, 1993).

There needs to be careful consideration of appropriate words and phrases in the translation process. Before translation of a questionnaire it is important that the meaning of key words and expressions are clearly understood, e.g. by providing descriptions of the ideas behind the key words and expressions, as well as the nature of the information expected to be obtained from the question(s). Such information can be of significant value to the translators.

Operational equivalence

The methods of collecting data, i.e. mode of administration, questionnaire format, instructions and measurement methods, will affect the results differently in the different cultures or language groups. Operational equivalence will be achieved when these different elements are shown not to affect the results.

Functional equivalence

Functional equivalence is the extent to which the questionnaire does what it is supposed to do, equally well across the different cultural/language groups. Assessing the extent to which functional equivalence has been achieved involves assessing the degree to which the other types of equivalence (described here) have been carried out. Only if 'reasonable equivalence' has been achieved in all of the areas described can it be argued that the results obtained are likely to be comparable.

The translation process

The forward-backward translation procedure (Brislin, 1970) is the one most commonly quoted in the adaptation and translation process. First, a forward translation is carried out using a bilingual person or persons who translate the questionnaire from language A to language B. The forward translation is then back-translated from language B to language A by a bilingual person or persons. Ideally, a panel of bilingual experts then compares the equivalence between the forward and backward translation. This procedure should continue until 'satisfactory equivalence' is

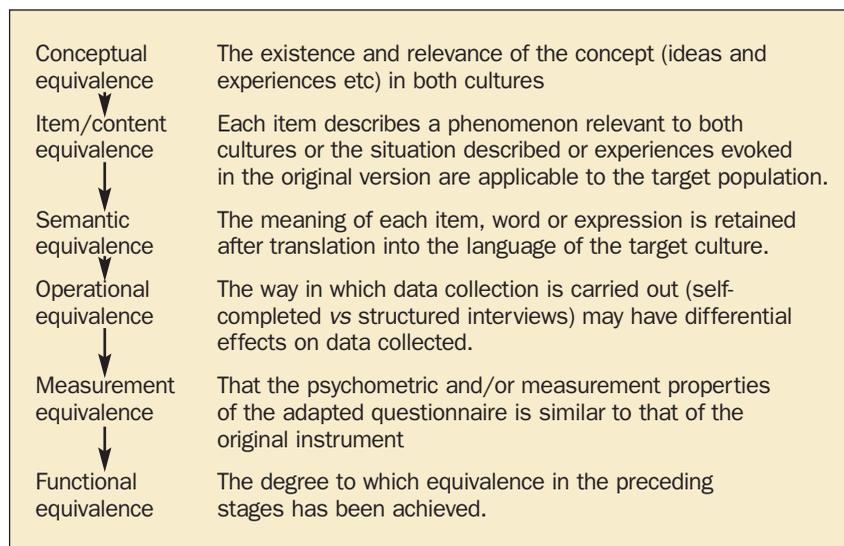
achieved between the original and translated versions.

When a questionnaire is to be translated, the following recommendations should be followed:

- Use linguistically competent translators who are conversant in the target languages
- Translators should be fully aware of the objectives of their role in the process and ideally have prior experience in adapting health-related questionnaires for cross-cultural use
- Some form of structured evaluation by the translators should be available, e.g. regarding difficulties experienced, which could enhance the ability to identify problems at an early stage of the process
- Forward translations should be produced by at least two independent translators to enable the identification of errors and misunderstandings resulting from the source version. The quality of the forward translation will be higher when teams rather than individuals carry out each translation
- As many back translations as forward translations should be produced
- A group of bilingual speakers should review and verify the equivalence between the source and final versions
- Pre-testing the translated questionnaire should be carried out on a representative sample of the population under study (Leplege and Verdier, 1995; Guillemin et al, 1993).

Despite all of these quality controls within the translation process, language remains a subjectively perceived form of communication. The objective of translation is to retain the intended meaning of the message, so there will always be an aspect of art involved in the process and in which the balance is achieved between rigidity, accuracy and the naturalness. For further reading on the cross-cultural adaptation of health-related questionnaires see

Figure 2. Types of cross-cultural equivalence of a questionnaire.



Bullinger et al, (1993); Meadows et al (1996); Bentzen et al, (1998); Meadows and Wisher (2000).

Conclusion

This article has attempted to describe the important aspects in the design of self-completion and interviewer administered questionnaires, as well as issues which need to be considered when adapting existing questionnaires, particularly for use in different cultural and language groups.

Constructing a questionnaire is not a simple process and requires skill, planning and care to ensure that all the relevant information is collected and that the information is both reliable and valid. Before undertaking the design of a questionnaire and using it, the novice researcher is strongly advised to discuss their research plans with a researcher experienced in questionnaire design so as to maximize the benefits of the process both in terms of the quality of the data collected and findings. ■

Bentzen N, Christianson T, McColl E, Meadows KA (1998) Selection and cross-cultural adaptation of health outcome measures. *Eur J Gen Pract* 4: 27–33

Bourque LB, Fielder EP (1995) *How to Conduct Self-Administered and Mail Surveys*. Sage, Thousand Oaks, CA

Brislin RW (1970) Back-translation for cross-cultural research. *J Cross Cultural Psych* 1: 185–216

Bullinger M, Anderson R, Cella D, Aaronson N (1993) Developing and evaluating cross-cultural instruments: From minimum to optimal models. *Qual Life Res* 2(6): 451–9

Dillman DA (1978) *Mail and Telephone Surveys: The Total Design Method*. Wiley, New York

Fowler FJ (1995) *Improving Survey Questions: Design and Evaluation*. Sage, Thousand Oaks, CA

Frey JH (1989) *Survey Research by Telephone*. 2nd edn. Sage, Newbury Park, CA

Frey JH, Oishi SM (1995) *How to Conduct Interviews by Telephone and in Person*. Sage, Thousand Oaks, CA

Guillemin F, Bombardier C, Beaton D (1993) Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *J Clin*

Epidemiol 46(12): 1417–32

Health Education Authority (1994) *Black and Minority Ethnic Groups in England*. Wheatons, Exeter

Herdman M, Fox-Rushby J, Badia X (1998) A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res* 7(4): 323–35

Hoinville G, Jowell R and associates (1978) *Survey Research Practice*. Heinmann Educational Books, London

Howlett B, Ahmad W, Murry R (1992) An exploration of white, Asian and Afro-Caribbean peoples concepts of health and illness causation. *New Community* 18: 281–92

Hsu FLK (1993) Psychosocial homogeneities and Zen: conceptual tools for advancing psychological anthropology. *American Anthropologist* 73: 23–44

Hunt S (1986) Cross-cultural issues in the use of socio-medical indicators. *Health Policy* 6(2): 149–58

Jenkins CR, Dillman DA (1997) Towards a theory of self-administered questionnaire design. In: Lyberg L, Biemer P, Collins M, de Leew E, Dippo C, Schwarz N et al eds. *Survey Measurement and Process Quality*. Wiley, New York: 165–96

Katz S, Ford AB, Moskowitz RW et al (1963) Studies of illness in age: The index of ADL, a standardized measure of biological and psychosocial function. *JAMA* 185: 914–19

Leplège A, Verdier A (1995) The adaptation of health status measures: methodological aspects of the translation procedure. In: Shumaker SA, Berzon R eds. *The International Assessment of Health-related Quality of Life: Theory, Translation, Measurement and Analysis*. Rapid Communications, Oxford: 93–101

Litwin M S (1995) *How to Measure Survey Reliability and Validity*. Sage, Thousand Oaks, CA

McColl E, Jacoby A, Thomas L et al (2001) Design and use of questionnaires: A review of best practice applicable to surveys of health service staff and patients. *Health Technol Assess* 5(31): 1–256

McColl E, Meadows K, Barofsky I (2003) Cognitive aspects of survey methodology and quality of life assessment. *Qual Life Res* 12(3): 217–18

Meadows KA (2003) So you want to do research? 3: An introduction to qualitative research methods. *Br J Community Nurs* 8(10): 464–9

Meadows K, Bentzen N, Touw-Otten F (1996) Cross-cultural issues: An outline of the important principles in establishing cross-cultural validity in health outcome assessment. In: Hutchinson A Bentzen N, König-Zahn, C, eds. *Cross-Cultural Health Outcome Assessment: A Users Guide*. The European Research Group on Health Outcomes. University of Hull

Meadows K, Wisher S (2000) Establishing cross-cultural validity in health surveys. National Centre for Social Research. *Survey Methods Newsletter* 20(2):1363–4518

Miller PV (1984) Alternative question forms for attitude scale questions in telephone interviews. *Public Opinion Quarterly* 48: 766–78

Morton-Williams J (1993) *Interviewer Approaches*. Dartmouth, Aldershot

Oppenheim AN (1992) *Questionnaire Design, Interviewing and Attitude Measurement*. 2nd edn. Pinter, London

Schuman H, Kalton G, Ludwig J (1983) Context and contiguity in survey questionnaires. *Public Opinion Quarterly* 47: 112–15

Schuman H, Ludwig J, Krosnick JA (1986) The perceived threat of nuclear war, salience and open questions. *Public Opinion Quarterly* 50: 519–36

Serdula MK, Mokdad AH, Pamuk ER et al (1995) Effects of question order on estimates of the prevalence of attempted weight loss. *Am J Epidemiol* 142(1): 64–7

Sirken M, Hermann D, Schechter S, Schwarz N (1999) In: Tanur J, Tourangeau R (eds). *Cognition and Survey Research*. Wiley, New York: 10

Sudman S, Bradburn N (1982) *Asking Questions: A Practical Guide to Questionnaire Design*. Jossey-Bass, San Francisco, CA

Touw-Otten F, Meadows K (1996) Cross-cultural issues in outcome measurement. In: Hutchinson A, McColl E, Cristie C, Riccalton C eds. *Health Outcome Measures in Primary and Out-patient Care*. Churchill Harwood Academic Publishers: 199–208

KEY POINTS

- Using a questionnaire to collect information for studies or surveys is now common practice.
- The mode of questionnaire administration is the first choice to be made.
- The wording and ordering of questions is central to their reliability and validity.
- A well-designed questionnaire can simplify the tasks of the respondent, interviewer and data processor as well as producing reliable data.
- Pre-testing and piloting the questionnaire is an integral part of its development.
- Adapting an existing questionnaire for a different purpose or group of people can have serious implications on its existing reliability and validity. Emphasis must be placed on achieving equivalence across the different cultural or language groups in the concepts measured and the items tapping these concepts.