Questionnaire Design

Introduction

The International Organization for Standardization (2012) points out that research findings can be affected by wording, question order, and other aspects of questionnaire design. The following guidelines present options for the deliberate design of questions intended for implementation in multinational, multicultural, or multiregional surveys, which we refer to as “3MC” surveys. In this context, “deliberate design” means that the questions have been specifically constructed or chosen for comparative research purposes, according to any of several criteria and strategies (Harkness, Edwards, Hansen, Miller, & Villar, 2010b). The models and strategies discussed here are applicable to a variety of disciplines, including the social and behavioral sciences, health research, and public opinion research.

This chapter presents a basic outline of the approaches available to develop questions for comparative studies, the procedures involved in each, and the advantages and disadvantages of the different approaches.

Although questionnaire design for 3MC surveys is related in various ways to question translation, adaptation, technical instrument design, pretesting, and harmonization, these topics are more fully addressed in other chapters (see Translation: Overview, Adaptation, Instrument Technical Design, and Pretesting).

This chapter borrows terminology from translation studies, which define “source language” as the language translated out of and “target language” as the language translated into. In like fashion, the chapter distinguishes between “source questionnaires” and “target questionnaires.” Source questionnaires are questionnaires used as a blueprint to produce other questionnaires, usually on the basis of translation into other languages (see Translation: Overview); target questionnaires are versions produced from the source questionnaire, usually on the basis of translation or translation and adaptation (see Adaptation). Target questionnaires enable researchers to study populations who could not be studied using the source questionnaire.

Guidelines

Goal: To maximize the comparability of survey questions across cultures and languages and reduce measurement error related to question design.
1. Ensure that questionnaire design follows basic best practice recommendations for general survey research.

**Rationale**

There are three general strategies for questionnaire design (Harkness, van de Vijver, & Johnson, 2003):

1.1 To re-use questions which seem suitable that have already been used in other surveys.
1.2 To adapt questions which have been developed for other purposes to suit new needs or populations.
1.3 To write entirely new questions.

Basic questionnaire design requirements need to be met regardless of which one of the three strategies is adopted and whether the project is comparative or not.

The procedural steps presented here identify fundamental aspects of questionnaire design with which researchers should be familiar before beginning work on any questionnaire and certainly before attempting comparative design. The steps do not provide guidance on each facet of design identified or on general design issues. A wealth of survey literature addresses these topics (e.g., see Bradburn, Sudman, & Wansink, 2004; Converse & Presser, 1986; Fowler, 1995; Groves, Fowler, Couper, Lepkowski, Singer, and Tourangeau, 2009; Willimack, Lyberg, Martin, Japec, & Whitridge, 2004).

**Procedural steps**

1.1 Review survey methods literature and research on basic aspects of general questionnaire design. Theories contributing to question/questionnaire design include:

1.1.1 Cognition and survey research, including theories of survey response (Schwarz, 1996; Tourangeau & Rasinski, 1988; Tourangeau, Rips, & Rasinski, 2000).
1.1.2 Measurement error and other sources of observational errors (Biemer & Lyberg, 2003; Groves, 1989).
1.1.3 Response styles and response bias (Baumgartner & Steenkamp, 2001; Johnson & van de Vijver, 2003; Schwarz, Oyserman, & Peytcheva, 2010; Vaerenbergh & Thomas, 2012; Yang, Harkness, Ching, & Villar, 2010).
1.1.4 Functional equivalence and comparability (Berry, 1969; Johnson, 1998a; Verba, 1969; Mohler and Johnson, 2010).

1.2 Review literature and research on the kinds of questions that can be asked (Bradburn et al., 2004; Converse & Presser, 1986; Dillman,
Cross-Cultural Survey Guidelines

Smyth, & Christian, 2009; Fowler, 1995; Groves, et al., 2009; Payne, 1980). Some of the kinds of questions listed below may overlap; for example, a factual judgment question may be about behavior or may ask for socio-demographic details.

1.2.1 Knowledge questions. Knowledge questions assess the respondent’s familiarity, awareness, or understanding of someone or something, such as facts, information, descriptions, or skills.

Example: Who is the President of the United States?

1.2.2 Factual judgment questions. Factual judgment questions require respondents to remember autobiographical events and use that information to make judgments (Tourangeau, et al., 2000). In principle, such information could be obtained by other means of observation, such as comparing survey data with administrative records, if such records exist. Factual judgment questions can be about a variety of things, such as figure-based facts (e.g., date, age, weight), events (e.g., pregnancy, marriage), and behaviors (e.g., smoking, media consumption).

Example: During the past two weeks, how many times did you see or talk to a medical doctor?

1.2.3 Socio-demographic questions. Socio-demographic questions typically ask about respondent characteristics such as age, marital status, income, employment status, and education. For discussion of their design and interpretation in the comparative context, see Granda, Wolf, & Hadorn (2010), Hoffmeyer-Zlotnik & Wolf (2003), and the International Organization for Standardization (2012). See also Translation: Overview and Adaptation.

Example: In what year and month were you born?

1.2.4 Behavioral questions. Behavioral questions ask people to report on things they do or have done.

Example: Have you ever smoked cigarettes?

1.2.5 Attitudinal questions. Attitudinal questions ask about respondents’ opinions, attitudes, beliefs, values, judgments, emotions, and perceptions. These cannot be measured by other means; we are dependent on respondents’ answers.

Example: Do you think smoking cigarettes is bad for the smoker’s health?

1.2.6 Intention questions on behavior. Intention questions ask respondents to indicate their intention regarding some behavior. They share features with attitudinal questions.

Example: Do you intend to stop smoking?

1.2.7 Expectation questions. Expectation questions ask about respondents’ expectation about the chances or probabilities that certain things will happen in the future. They are used in
several cross-national surveys, such as Survey of Health, Ageing and Retirement in Europe (SHARE).

Example: *Thinking about the next ten years, what are the chances that you will receive any inheritance, including property and other valuables?*

1.3 Review literature and research on question formats.

1.3.1 **Closed-ended question** format. In closed-ended question formats, the survey question provides a limited set of predefined answer categories from which respondents choose.

Example: *Do you smoke?*

Yes ___

No ___

1.3.2 **Open-ended question** format. Open-ended question formats require respondents to answer questions in their own words.

Example: *What is your occupation?*

*(Please write in the name or title of your occupation)*

1.4 Review literature and research on response scales. Response scales are predefined sets of answer categories for a closed question from which respondents are asked to select a response. Common response scales include **rating**, **ranking**, and **frequency** scales.

1.4.1 Rating uses an ordered scale of response options and requires the respondent to select one position on the scale.

Example: *To what extent do you agree or disagree with the following statement? It is a good idea to ban smoking in public places.*

Strongly agree

Somewhat agree

Neither agree nor disagree

Somewhat disagree

Strongly disagree

1.4.2 Ranking is a response format where respondents express their preferences by ordering persons, brands, etc., from top to bottom, generating a rank order of a list of items or entities. Ranking can be partial, where a longer list of responses is presented, and respondents are requested to rank a limited number.

Example: *Listed below are possible disadvantages related to smoking cigarettes. Please enter the number 1, 2, 3, or 4 alongside each possible disadvantage to indicate your rank ordering of these. 1 stands for the greatest disadvantage, 4 for the least disadvantage.*

___ Harmful effects on other people’s health

___ Stale smoke smell in clothes and furnishings
___ Expense of buying cigarettes
___ Harmful effects on smoker’s health

Example of partial ranking: Out of these 13 qualities in children, please rank the 5 qualities you think are most desirable in children (Kohn, 1969).
___ Has good manners
___ Tries hard to succeed
___ Is interested in how and why things happen
___ etc.

Card sorts are another ranking technique, wherein words, statements, graphics, etc., are written onto cards which the respondent arranges according to some dimension.

1.4.3 Frequency scales are a response format where respondents are required to select the option that best describes the frequency in which certain behaviors occur.

Example: How often did you attend live music events in past year?

Never
Rarely
Sometimes
Often
Always

1.5 Review literature and research on types of data. Data from rating, ranking, and frequency scales can be categorized as nominal, ordinal, and numeric (Fink, 2003).

1.5.1 Nominal data. Data are nominal or categorical when respondents are asked to name, or categorize, their answer. In nominal data, there is no numeric way to rate, rank, or otherwise differentiate response categories.

Example: Which of these political parties did you vote for in the last national election?

Republican Party
Democratic Party
Socialist Party
Libertarian Party

1.5.2 Ordinal data. Data are ordinal when respondents are asked to rate or order items on a list. In ordinal data, there are no real numeric values associated with the categories, and there is no way to measure distance between categories. However, the categories can be ranked or rated from one end of a scale to the other.

Example: How much do you disagree or agree with the statement: Democracy is the best form of government.

Strongly agree
Agree
Disagree
Strongly disagree

1.5.3 Numeric data. There are two types of numeric data: interval data and ratio data, although statistically these two types of data tend to be treated the same. With interval data, the distances between the response values (i.e., numbers) have real meaning.

Example: The Fahrenheit temperature scale, where a 10-point difference between 70°F and 80°F is the same as a 10-point difference between 40°F and 50°F.

Measurements of physical properties have characteristics whose quantity or magnitude can be measured using ratio scales. Ratio measurements have a true zero, unlike interval data, and comparisons between data points can be made accordingly.

Example: The Kelvin temperature scale, where 50 kelvins is half as warm as 100 kelvins.

Other examples include time (measured in seconds, minutes, hours), meters, kilograms.

1.6 Review literature and research on mode (i.e., the means by which data are collected) (de Leeuw, 2008; Dillman, et al., 2009; Smith, 2005). The choice of mode will affect the design options for various aspects of questionnaire and survey instrument design (e.g., length of the questionnaire, layout of instruments, and application of visual stimuli). (See Study Design and Organizational Structure and Instrument Technical Design.)

1.6.1 In terms of the standard literature, “mode” is related to whether an interviewer enters the data (as in telephone and face-to-face interviews) or the respondent enters the data (as in web surveys and paper-and-pencil surveys).

1.6.2 A second relevant aspect is the channel of communication (visual, oral, aural, tactile).

1.6.3 A third is the sense of privacy. Usually, self-administered survey modes create a greater sense of privacy than interviewer administered modes.

1.7 Review literature on techniques can be used in survey questionnaire design. Random-response technique (RRT) is a method designed to elicit reliable responses to sensitive survey items, although it is only useful for a very limited number of yes/no questions in any given survey. In RRT, respondents are randomly assigned to answer one of two yes/no questions: one sensitive and the other non-sensitive and with a known probability. The interviewer is unaware of which question is given to the respondent, and only records the answer. Based on the probability of selecting the sensitive question, the
probability of respondents who answer yes to the nonsensitive question, and the proportion of respondents who answer yes to the RRT question, the is technique can be used to calculate the proportion of respondents who give an affirmative answer to the sensitive question. In a study of abortion behavior among Mexican women, RRT generated the most reliable data regarding induced abortion, when compared to face-to-face surveys and both audio and paper-based self-administered modes (Lara, Strickler, Olavarrieta, & Ellertson, 2004).

2. Become familiar with the comparative design options available and the advantages and disadvantages of each.

**Rationale**

Knowledge of the different approaches available for comparative design for surveys in multiple cultures, languages, or countries enables researchers to make informed choices for their projects.

**Procedural steps**

2.1 Read relevant literature (and, if possible, talk to primary researchers) to become familiar with the advantages and disadvantages of the major approaches to questionnaire design for 3MC surveys. The three basic approaches involve asking the same questions and translating (ASQT), asking different questions (ADQ, usually to adapt to new cultural, social or other needs), or using a mixed approach that combines ASQT and ADQ (Harkness, 2008b; Harkness, et al., 2010b; Harkness et al., 2003). See Translation: Overview for more information on methods of translation, and Adaptation for more examples of ADQ.

2.1.1 **Ask the same questions and translate (ASQT).** In this approach to question design, researchers ask a common set of questions of all populations studied.

- The most common way to do this is by developing a source questionnaire in one language and then producing other language versions, usually on the basis of translation or translation and adaptation. A TRAPD (Translation, Review, Adjudication, Pretesting, and Documentation) team translation model is suggested. See Translation: Overview for more information.
- **Decentering** is a second way to “ask the same questions.” With decentering, the same questions are developed simultaneously in two languages -- there is no source questionnaire or target language questionnaire. The decentering process removes culture-specific elements
from both versions (Harkness, 2008b). However, decentering is only suitable for two language projects and its use is restricted in many ways (Harkness, 2008b; Harkness, et al., 2010b; Werner & Campbell, 1970). It is also very work intensive and there is little information about recent experiences using this technique. See Harkness, et al. (2010b) for a more detailed explanation of decentering.

- The key advantage of the ASQT approach is standardization of the stimuli across cultures; the main disadvantage of the ASQT approach is that a literal or near close translation may not be culturally suitable and appropriate or may not be possible. For example, the European Social Survey (ESS) had no issue using ASQT to translate, “Do you consider yourself as belong to any particular religion or denomination?” into multiple languages; however, ASQT was not appropriate for, “Do you have difficulty walking several blocks?” Using ASQT for the latter was judged problematic in study countries where neighborhoods are not organized into blocks.

- **Anchoring vignettes** allow researchers to make adjustments when respondents from different cultures, countries, or ethnic groups interpret questions in different ways (King, Murray, Salomon, & Tandon, 2004). Participants are asked to provide assessments both for themselves and for several hypothetical people. Anchoring vignettes assume vignette equivalence (i.e., the hypothetical situations portrayed in the vignettes are viewed equivalent across cultures to be compared) and reporting consistency (i.e., respondents rate their own and hypothetical vignette persons in a consistent way). They can be analyzed by both nonparametric methods and model-based parametric methods. However, the profile of hypothetical persons in the vignettes can potentially affect the adjustments (Grol-Prokopczyk, 2014). Methods have been developed for the evaluation and selection of anchoring vignettes for a diverse range of topics (King & Wand, 2007). See King (n.d.) for more information and resources.

2.1.2 **Ask different questions (ADQ).** In this approach, researchers ask the most salient questions for each population on a given common construct or conceptual domain. The different questions and, possibly, different indicators used in each location are assumed to tap a construct that is shared across populations. For example, the following questions may all be effective indicators of the concept of intelligence for individual populations. However, characteristics of intelligence may be
more or less salient, depending on local context and ADQ may be the best strategy (Harkness, 2008b). See Adaptation for more information.

Is she quick-witted?
Does she give considered responses?
Is she good at knowing whom to ask for help?
Is she good at finding solutions to urgent problems?

- This approach emphasizes the standardization of meanings and strives for functional equivalence.
- The downside of this approach is that the item-by-item analyses across populations are more difficult to justify since the questions are not the same across different groups.

2.1.3 A mixed approach that combines ASQT and ADQ. Many 3MC surveys use a mix of ASQT and ADQ questions.

- Some questions blend a common part (ASQT) and country-specific parts (ADQ). Socio-demographic questions on education, for example, are often asked in terms of a shared question stem (such as “What is the highest level of education you have completed?”), accompanied by local/national categories of educational level or qualification (ADQ). These are then mapped onto an international standard (see Translation: Harmonization).

- For cross-cultural surveys, cultural adaptation of instruments along with translation improves measurement comparability (Georgas, Weiss, van de Vijver, & Saklofske, 2003; Hambleton, Merenda, & Spielberger, 2005). See the Adaptation chapter for more details.

2.2 Weigh the advantages and disadvantages of each approach in terms of the study design (see overview in Appendix A).

2.3 Decide on the most viable approach for the study within a quality framework that addresses survey error related to questionnaire design (see Survey Quality).

2.4 Match the question style (responses) to respondent recall style. For example, incorporate calendar techniques (e.g., event history calendars or life history calendars; see Data Collection: General Considerations) for people who identify time by events (Yount & Gittelsohn, 2008). Of course, this requires researchers with substantive knowledge of each cultural group in the survey.
Lessons learned

2.1 Not all options will be available for every study. The study design, the target population, and the mode required may all impose constraints. For example, if questions from other studies are to be used again ("replicated"), only an ASQT model (perhaps with adaptation) is possible for these questions. The chosen data collection method, the sample design, the fielding schedules, and available funds or stipulations on the use of funds can all limit options (see Study Design and Organizational Structure, Data Collection: General Considerations, Sample Design, and Tenders, Bids, and Contracts).

2.2 Cross-cultural questionnaire design literature can be hard to locate, unclear, or very sparse on details. Even detailed study reports might be clear to people involved in a project but not clear enough for "outside" readers. Detailed and transparent documentation of the questionnaire design process is critical for cross-cultural survey research in order for other data users to understand the data collection procedures in each country and to evaluate the data quality in a comparative manner.

2.3 Researchers are usually most familiar with the ASQT approach, but may not be aware of the limitations and constraints of this approach (Behr, 2010; Harkness, et al., 2010b; Harkness et al., 2003; Harkness, Villar, & Edwards, 2010a). In addition, pressures to replicate questions might over-promote the ASQT approach. Please see Appendix A for the pros and cons of ASQT.

2.4 Comparability or equivalence is sometimes judged on the basis of similar wording across questionnaires. This is, indeed, what is often targeted in ASQT approaches. However, even nominally “accurate” translations do not necessarily produce comparable data (see Translation: Overview). For example, a close translation of the English question “Does he like adventures?” in French is more likely to be understood as “Does he like amorous adventures?” Bilingual or multi-lingual researchers with substantive knowledge of two or more cultures and languages are essential in this approach. In addition, qualitative study and cognitive testing are critical for questionnaire translations. After all, the mutual understanding among the respondents is the goal.

2.5 It is difficult to find examples of surveys with most substantive questions based on an ADQ approach. There are examples of research that analyzes different questions from different studies and takes them to reflect aspects of a given common construct (Van Deth, 1998).
2.6 Change of question formats or adapted questions can radically affect respondents’ answers in cross-cultural surveys, such as in the International Social Survey Programme (ISSP) (Smith, 1995) (see also Adaptation).

2.7 Researchers also need to be aware of the negative consequences associated with inappropriate standardization (see Harkness & Behr, 2008, and Lynn, Japec, & Lyberg, 2006).

2.8 Researchers need to be aware of cross-cultural differences in the relevance, saliency, and social desirability of survey questions. For instance, certain questions may not be relevant or salient for a given population and that population may not have the information necessary to answer those questions. In addition, questions considered innocuous in one culture may be threatening or taboo in other cultures. For example, respondents in Islamic countries would find questions about alcohol use or the number of children born to an unmarried respondent offensive (e.g., see Smith, 2002).

2.9 Respondents’ social reality and cultural framework shape their perceptions and survey responses in a variety of ways (see Braun & Mohler, 2003, and Yang, et al., 2010).

3. Establish a lead team or working group responsible for questionnaire design, and appoint a coordinator responsible for organizing scheduling, communication channels and rules, and the design deliverables.

Rationale

Good questionnaires can rarely be developed by a single person. This is especially true for 3MC research. In accordance with a quality assurance framework for design, a team is needed that provides the spread of knowledge, diverse skills, and cultural backgrounds for which successful comparative design calls (Lyberg & Stukel, 2010).

Procedural Steps

3.1 Decide, as appropriate, on the lingua franca and communication mediums to be used in the overall project and in the work of the questionnaire design team.

3.2 Identify a lead person in the design team who is also responsible for coordinating with other groups in the project (such as the coordinating center, if one exists – see Study Design and Organizational Structure).
3.3 Decide on appropriate communication channels (e.g., in-person and telephone meeting, or video-conferencing, including online meetings). Meet regularly to communicate progress.

3.4 Identify the various skills required in the team.
   3.4.1 These include all the skills needed for questionnaire design in general, including but not limited to 3MC research.
   3.4.2 They also include special expertise or skills relevant for designing a comparative instrument (e.g., understanding design models such as ASQT and ADQ), understanding the cultural impact of conceptual coverage, cultural norms that affect common ground and response processes, response styles, local population structure and needs, etc.).
   3.4.3 Depending on their roles in the team, members may need to be conversant in whatever lingua franca is used in a multilingual project.

3.5 Ensure that the team members recruited are from a reasonable spread of the countries, locations, or cultures participating in the study.

3.6 Ensure that the members recruited for the questionnaire design team have the skills and abilities needed for good questionnaire design. A questionnaire design team should consist of 1) comparativists, including area/cultural specialists, 2) substantive/subject area experts, 3) linguistic experts and 4) survey research experts (Mohler, 2006).
   3.6.1 If the cultural and linguistic experts in the project lack fundamental knowledge in survey research, it is important to provide training to them or include survey methodologists on the team.

3.7 If qualitative components are included, involve an interdisciplinary decision-making team with training in both qualitative and quantitative methods (Massey, 1987).

3.8 Identify the responsibilities of each member at an appropriate level of detail.

3.9 Recruit collaborators and external experts, as necessary and feasible, from the different populations involved. This ensures the availability of expertise on given topics and local knowledge. A drafting team might need specific and short-term input from an expert on a substantive area in the questionnaire. For example, if input on pensions is needed, an expert on the topic may be brought in exclusively for the development of pension-specific questions.
Lessons learned

3.1 In addition to the lead team, each cultural group can benefit from strong input from local participants who are similar to the intended sample population. Ways should be found to have any groups who are participating in the project, but are not directly part of the core development team, to contribute to the development of the questionnaire. This can be a less formal team of local participants who can help guide questionnaire development from the ground up, along the lines of “simultaneous [questionnaire] development” (Harkness, et al., 2010b; Harkness et al., 2003). Another option is for the working group to specify target variables while allowing local participants to specify the particular questions (Granda, et al., 2010). It will be helpful for local participants to be familiar with survey research methods.

3.2 Qualitative methods such as focus groups and cognitive interviews can be used to gain insights into the local community and experiences of the target population, which researchers alone may not be able to recognize or capture. Findings from qualitative methods can be used to inform questionnaire design and subsequent interpretation of quantitative results (Habashi & Worley, 2009).

4. Establish the procedures and protocols for questionnaire development and for testing at different stages in this development.

Rationale

Clear identification of the procedures and the protocols to be followed is essential to inform all those involved and to effectively implement and assess the chosen design process.

While different studies follow different design models (ASQT, ADQ, mixed approaches), this guideline identifies some of the key generic elements to be considered.

Procedural steps

4.1 Establish which design and related procedures are to be used (e.g., ASQT source questionnaire and translation).

4.2 Develop the protocols relevant for the chosen design model and the processes it calls for (e.g., protocol for questionnaire development of a source questionnaire intended for use in multiple locations and cultures/languages).
4.3 Create a schedule and budget for the milestones, deliverables, and procedures involved in the chosen design model. In the ASQT model this would include schedules and a budget for producing draft source questionnaires, review by participating cultures or groups, deadlines for feedback, schedules for pretesting, schedules for language harmonization, schedules for translation (See Translation: Scheduling), and subsequent assessment and pretesting. The participation of team members from all countries throughout the process is essential in ensuring that the questions are developed, translated, and tested appropriately among all target populations.

4.4 Create a framework of quality assurance and quality control to ensure compliance with protocols and the adequacy of outputs (see Survey Quality).

4.5 Create communication channels and encouragements which ensure that participants can and do make feedback on draft designs they are asked to review.

Lessons learned

4.1 Not all participating groups in a project will be confident that their input in the developmental process is (a) valuable in generic terms for the entire project, (b) accurate or justified, and (c) welcomed by perceived leading figures or countries in either the design team or the larger project. It is important to make clear to participating groups that every contribution is valuable. Sharing feedback across the project underscores the value of every contribution and explains to participating groups why their suggestions are or are not incorporated in design modifications.

5. **Pretest** source and target questionnaires.

**Rationale**

Questionnaires need to be pretested before they are used. The source questionnaire needs to be assessed for its suitability as a source questionnaire for multiple other versions, rather than as a questionnaire for a single population. Pretesting often relies on expert review, particularly for reviewing the suitability of the source questionnaire in other cultural groups.

The other versions produced—most likely on the basis of translation or translation and adaptation—also need to be pretested for suitability, ideally with every target population in the study. Various qualitative and quantitative approaches can be taken to pretest the target questionnaires.
in the target population. An often-cited recommendation is “If you do not have the resources to pilot-test your questionnaire, don't do the study.” ([Sudman & Bradburn, 1982], p. 283).

**Procedural steps**

(For detailed information about pretesting, see [Pretesting](#).)

**Lessons learned**

5.1 Pretesting is essential. Even questions previously used in other questionnaires must be tested for their suitability in a new context and for use with new populations.

5.2 Where possible, pretesting of the source questionnaire should be combined with pretesting a spread of other languages representing the diverse target populations in the project ([Skevington, 2002](#)).

5.3 Ensuring the quality of questionnaire development prior to pretesting is just as important as pretesting itself. Proper team selection, adequate briefing on requirements and expectations, and good use of documentation will enhance the quality of the questions presented for pretesting so that pretesting serves the monitoring and refining purposes it should have.

5.4 Combine both quantitative and qualitative techniques to evaluate and test questionnaires.

5.4.1 Question design and statistical modeling “should work in tandem for survey research to progress” ([Presser, et al., 2004](#)). In other words, when designing questions, consider how they will be used in analysis.

5.5 Even locations sharing the language of the source questionnaire (e.g., the U.S. and the U.K.) need to review the instrument for local suitability ([Jowell, 1998](#)).

6. **Establish a quality assurance and quality monitoring framework for questionnaire development.**

**Rationale**

Irrespective of the design approach followed to produce a questionnaire, quality standards must be set. These are critical to establishing quality assurance and quality monitoring steps for the process of developing any questionnaire ([International Organization for Standardization, 2012](#)).
**Procedural steps**

6.1 Be cognizant of possible sources of survey error in the questionnaire design phase and develop quality assurance and quality monitoring steps to address these (see *Survey Quality*). Possible sources of error in this phase include *validity* and measurement issues (*Groves et al., 2009*).

6.2 Acquaint question designers with important quality assurance literature on the topic of question design (e.g., on validity, tests of conceptual coverage, response process, sources of *measurement error*) (*Biemer & Lyberg, 2003; Groves et al., 2009*).

6.3 For source questionnaires, form a team in each country or location that meets to discuss the development and assessment of the source questionnaire at each phase. The team should have, or should be provided with, the methodological expertise needed for this task.

6.4 Have such teams document and report any queries or problems to the questionnaire drafting group in a timely fashion during the development phases or, as appropriate, report to the coordinating center (*Lyberg & Stukel, 2010*).

**Lessons learned**

6.1 Quality assurance and quality monitoring should be addressed early in the design planning process. *Thornton, et al., (2010)* describe the explicit procedures and protocols he and his research team followed in designing a multinational study with no pre-existing survey items to measure underlying theoretical concepts of the research question. Because error due to measurement and validity of brand-new measures were of particular concern, all survey question writing (and subsequent translation) was done as a team through a series of weekly meetings, beginning at the inception of the research project. All meetings involved collaborators, experts in a wide variety of relevant disciplines, from each study country. Meeting discussions, item wording decisions, and inconsistent field results pointing to measurement issues were carefully documented and necessary deviations from between-country comparability were detailed in the dataset codebook for future users (*Thornton et al., 2010*). See also *de Jong & Young-DeMarco (2017)* for a similar discussion of protocols followed for a cross-national comparative survey in the Middle East.

6.2 Variations in country-level assessment experience, research traditions, and methodological rigor regarding question design need
to be thoroughly investigated and understood when setting quality standards. Some locations or countries will need more assistance than others in understanding the relevance of some requirements. They may also need guidance on how products can be assessed in terms of these requirements.

6.3 Some entity, such as a questionnaire drafting group coordinator or a coordinating center, must be appointed to lead on these matters.

6.4 Through their knowledge of their own location and culture, local level collaborators and team members may well provide insights that other team members lack, even if quite experienced in questionnaire design.

7. Develop qualitative and quantitative protocols and procedures for assessing the quality of questions across survey implementations.

Rationale

Identifying standards to be met and establishing the criteria required to meet them, as well as agreeing on the good/best practice procedures to follow, are basic to undertaking quality assurance and quality monitoring.

Procedural steps

7.1 Determine appropriate methods to assess the quality of questions. Consider question standards and survey determinants (e.g., funding and resources), as well as the model of design chosen for the topic (Groves et al., 2009).

7.2 Include qualitative and quantitative methods of assessment (see the Pretesting chapter for a detailed description of assessment methods).

7.2.1 Qualitative options include:

- Various pretesting techniques, such as focus groups and cognitive interviews (see Pretesting).
- Expert appraisals by such groups as target population members, substantive experts, question design experts, or translators.
- Debriefings from any testing (interviewers and respondents).

7.2.2 Quantitative methods of assessment include pilot studies (Biemer & Lyberg, 2003; van de Vijver & Leung, 1997):

- Reliability (e.g., Cronbach’s alpha) and validity.
- Exploratory and confirmatory analyses such as variance analysis, factor analysis, multi-trait multi-method, item
response theory, latent class analysis, differential item functioning, or stand-alone or embedded experiments.

- These methods require planning as they require specific types of data.

7.3 When possible, use wording experiments to decide between different candidate question wordings (Fowler, 2004; Moore, Pascale, Doyle, Chan, & Klein Griffiths, 2004). However, little effort has been devoted to comparative experimental study on survey design and translation (for examples, see Harkness, Villar, Kephart, Schoua-Glusberg, & Behr, 2009).

7.4 Consider using advance translations or translatability assessment as part of questionnaire design to minimize later translation problems.

Lessons learned

7.1 Both qualitative and quantitative methods of assessment are necessary. Reliance on one without the other is not advised.

7.2 Do not use pretesting as the main tool for question refinement. Make the questions as well designed as possible before pretesting so that pretesting can be used to find problems not identifiable through other refinement procedures.

7.3 Different disciplines favor and can use different developmental and testing procedures, partly because of their different typical design formats. Social science surveys, for example, often have only one or two questions on a particular domain; psychological and educational scales, on the other hand, might have more than twenty questions on one domain.

8. Develop a documentation scheme for questionnaire design decisions, design implementation, and quality assurance protocols.

Rationale

Documentation aids in producing the questionnaire and can be a tool for quality assurance and monitoring. As indicated in Survey Quality, continual measurement and documentation of the quality targeted and achieved is necessary to identify quality problems. Even if sources of error are not recognized until later, documentation can be used to inform improved designs for future studies.
Procedural steps

8.1 Design the documentation process before question development begins and document question design from the start. This ensures that all decisions are captured and that action can be taken in a timely fashion.

8.2 Standardize documentation requirements and formats across all locations or countries involved in question development. This facilitates feedback in an ASQT model and comparable development in an ADQ model.

8.3 Create flexible documentation templates that allow categories to be added if unforeseen issues arise.

8.4 Create a clear and concise description of the questionnaire design procedures which is user-oriented and user-friendly. Include:
   8.4.1 Conceptualization from concept to questions.
   8.4.2 Operationalization (approach, mode, development across versions, adaptation agreements, annotations, (shared) language harmonization, origin of questions whether new, replicated, adopted, or adapted).
   8.4.3 An analysis plan.

8.5 Record the development of indicators and questions from start to finish (e.g., any modifications made to questions at different stages and why).

8.6 Version control procedures are necessary whenever a source questionnaire is modified across time.
   8.6.1 A version of the source questionnaire will serve as the gold standard, or source version 1. Document any changes made to it over time.

Lessons learned

8.1 Documentation must accompany questionnaire design since it will be used to detect problems in time to address them.

8.2 If documentation is left to the end of questionnaire design (or even later), details will be forgotten and intervention will not be possible. Study monitoring questionnaires for the ISSP (completed well after question design and translation have been completed) sometimes contain documentation on translation challenges for two or three phrases. The templates used in recent German ISSP translation discussions note a myriad of challenges (Behr, 2010).
8.3 Any changes countries make to their design protocols and procedures and any reservations they have about development must be carefully documented. If these are made available in a timely fashion to either the questionnaire drafting coordinator or, as appropriate, the coordinating center, problems can be addressed. For example, feedback to questionnaire drafting groups from countries participating in the ISSP and ESS studies sometimes lead to changes in draft versions of source questions.

8.4 At a later stage, documentation might be helpful in understanding potential differences in the data, either over the course of the study (within a country) or across variables (between countries).

8.5 Providing tools to make the job easier encourages people to engage in the task and ensures better documentation.

8.6 Demonstrating the importance of documentation motivates people to engage in it. Even simple things can help convince and motivate — for example, showing how a template can help check for flipped order of answer categories across a range of questions.
## Appendix A

### Some advantages and constraints on different approaches to question design

<table>
<thead>
<tr>
<th>Approach</th>
<th>Advantages</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ask the same question and translate (ASQT)</td>
<td>If successful, questions and item scales can be compared one-by-one across data sets and thus permit the most sophisticated analyses based on what is sometimes called full scalar equivalence (see van de Vijver &amp; Leung, 1997).</td>
<td>Developing for an ASQT questionnaire may result in reduced specificity of questions used and a resultant loss of saliency and fine-grained information.</td>
</tr>
<tr>
<td>ASQT is the least complicated approach to organize and implement. This is not to suggest it does not involve considerable effort as reflected in this chapter and Translation: Overview.</td>
<td></td>
<td>Conceptual coverage for all or some of the populations in the study may thus be reduced and not comparable across populations.</td>
</tr>
<tr>
<td>Researchers engaged in it and clients requesting or expecting it may feel more like they are on familiar territory with this model than with others.</td>
<td>At the translation stage, those implementing the ASQT may have inappropriate goals for the translation product and produce poor target language versions.</td>
<td></td>
</tr>
<tr>
<td>ASQT potentially permits replication of existing questions—provided their basic suitability for translation and fielding with the target populations is ensured.</td>
<td>Replicated questions encourage close translation and may not be optimal for one or more target populations.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ASQT does not work well at all for some kinds of questions (e.g., background variables such as education).</td>
</tr>
<tr>
<td>Approach</td>
<td>Advantages</td>
<td>Constraints</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ASQT and adapt approaches</td>
<td>Call for expertise in question development and translation in areas still requiring basic research</td>
<td>and/or training.</td>
</tr>
<tr>
<td>Decentering</td>
<td>Allows two questionnaires to be developed in collaboration and creates the potential for full scalar</td>
<td>May result in questions with low saliency for either culture since anything</td>
</tr>
<tr>
<td></td>
<td>equivalence (see van de Vijver &amp; Leung, 1997).</td>
<td>that constitutes a problem in the course of development is removed or altered.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>This would be an obstacle to full scalar equivalence.</td>
</tr>
<tr>
<td></td>
<td>Avoids the situation where the needs of one questionnaire and language/culture dominate.</td>
<td>Decentering is not viable for projects involving more than a handful of</td>
</tr>
<tr>
<td></td>
<td></td>
<td>languages.</td>
</tr>
<tr>
<td></td>
<td>Can be useful in developing comparable questions for very disparate cultures.</td>
<td>Decentering is very work intensive and there is little information about</td>
</tr>
<tr>
<td></td>
<td></td>
<td>recent experiences using this technique.</td>
</tr>
<tr>
<td>Ask different questions</td>
<td>Avoids the need to base questionnaires for various cultures and languages on translation of a source</td>
<td>Little detailed information is available about recent projects adopting an</td>
</tr>
<tr>
<td>approaches</td>
<td>questionnaire.</td>
<td>ADQ approach.</td>
</tr>
<tr>
<td></td>
<td>Researchers can select the indicators and questions considered most salient for a given population</td>
<td>Researchers have few guidelines about how to develop the quality assurance</td>
</tr>
<tr>
<td></td>
<td>provided these produce data which can still be compared across populations.</td>
<td>and quality control steps needed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>If different populations are only asked questions developed for them, item-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>by-item analyses across populations are more difficult to justify.</td>
</tr>
<tr>
<td>Approach</td>
<td>Advantages</td>
<td>Constraints</td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td>It is easier for a group joining an ADQ-based study after other groups have developed and fielded their questionnaires to produce a suitable questionnaire for their context than it is for researchers joining an ASQT project after the source questionnaire has been finalized.</td>
<td>Most researchers and clients are unfamiliar with ADQ approaches.</td>
</tr>
<tr>
<td>Mixed approaches combining ASQT and ADQ components</td>
<td>These can combine the advantages of ASQT and ADQ.</td>
<td>They increase the number and kind of procedural steps to be implemented and assessed.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>They call for expertise in areas still requiring basic methodological research.</td>
</tr>
</tbody>
</table>
References


Lyberg, L. E. & Stukel, D. M. (2010). Quality assurance and quality control in


Skevington, S. M. (2002). Advancing cross-cultural research on quality of life:
Observations drawn from the WHOQOL development. *Quality of Life Research, 11*, 135-144.


Further Reading


