*Janet Harkness with (alphabetically) Ipek Bilgen, AnaLucía Córdova Cazar, Mengyao Hu, Lei Huang, Sunghee Lee Mingnan Liu, Debbie Miller, Mathew Stange, Ana Villar, and Ting Yan, 2016*

**Introduction**
**Guidelines**
**1. Ensure that questionnaire design follows basic best practice recommendations for general survey resear**
**2. Become familiar with the comparative design options available and the advantages and disadvantages o**
**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.**
**4. Establish the procedures and protocols for questionnaire development and for testing at different stages development.**
**5. Pretest source and target questionnaires.**
**6. Establish a quality assurance and quality monitoring framework for questionnaire development.**
**7. Develop qualitative and quantitative protocols and procedures for assessing the quality of questions acr survey implementations.**
**8. Develop a documentation scheme for questionnaire design decisions, design implementation, and quality assurance protocols.**
**References**
**Appendices: A**
**Further Reading**

---

## Introduction

The points out that research findings can be affected by wording, question order, and other aspects of questionnaire The following guidelines present options for the deliberate design of questions intended for implementation in multi multicultural, or multiregional surveys, which we refer to as '3MC' surveys. In this context, “deliberate design” mea the questions have been specifically constructed or chosen for comparative research purposes, according to any of s criteria and strategies . The models and strategies discussed here are applicable to a variety of disciplines, including 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, techni instrument design, pretesting, and harmonization, these topics are more fully addressed in other chapters (see Transl Overview, Adaptation, Instrument Technical Design, and Pretesting).

This chapter borrows terminology from translation studies, which define 'source language' as the language translate 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 o questionnaires, usually on the basis of translation into other languages (see Translation: Overview); target questionn versions produced from the source questionnaire, usually on the basis of translation or translation and adaptation (se Adaptation). Target questionnaires enable researchers to study populations who could not be studied using the sourc questionnaire.

↑ Ba

## Guidelines

**Goal**: To maximize the comparability of survey questions across cultures and languages and reduce measurement er related to question design.

↑ Ba

**1. Ensure that questionnaire design follows best practice recommendations for both general and 3MC survey research.** ***Rationale***

There are three general strategies for questionnaire design :

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

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

The procedural steps presented here identify fundamental aspects of questionnaire design with which researchers sh familiar before beginning work on any questionnaire and certainly before attempting comparative design. The steps overview, but cannot provide exhaustive guidance on each facet of design identified or on all general design issues. wealth of survey literature addresses these topics (see , , , , and ).

It is important to note that any existing design recommendation relating to different types of questions (e.g., behavic attitudes, knowledge) or survey mode (e.g., face to face, telephone, Web), for example, must be carefully considered appropriateness to the culture and language of each population . When designing the questionnaire and other survey materials, researchers must attempt to identify and be informed by ways in which members of different cultures may systematically, in how questions are understood and answered, and in how culture may affect or interact with key questionnaire design features (such as response scales or question order) and topics. Understanding of the populatio interest and thorough pretesting are essential for the identification of potential problems with questions and element questionnaire design to avoid measurement and nonresponse error and threats to comparability in the results.

***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 .

1.1.2 Measurement error and other sources of observational errors .

1.2 Review literature and research on the kinds of questions that can be asked . Some of the kinds of questions 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** 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** require respondents to remember autobiographical events and use that information to make judgments . In principle, such information could be obtained by other means of obs such as comparing survey data with administrative records, if such records exist. Factual judgment quest 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** typically ask about respondent characteristics such as age, marital statu income, employment status, and education. For discussion of their design and interpretation in the comp context, see , , and the . See also Translation: Overview and Adaptation.

- Example: *In what year and month were you born?*

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

- Example: *Have you ever smoked cigarettes?*

1.2.5 **Attitudinal questions** ask about respondents' opinions, attitudes, beliefs, values, judgments, emotions 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** ask respondents to indicate their intention regarding some behavior. They share f with attitudinal questions.

- Example: *Do you intend to stop smoking?*

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

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

1.3 Review literature and research on question formats.

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

- Example: *Do you smoke?*

*Yes* ___
*No* ___

1.3.2 **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 closed question from which respondents are asked to select a response. Common response scales include ratin ranking, and frequency scales.

1.4.1 **Rating** uses an ordered scale of response options and requires the respondent to select one position on 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, brand from top to bottom, generating a rank order of a list of items or entities. Ranking can be partial, where a 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 n 2, 3, or 4 alongside each possible disadvantage to indicate your rank ordering of these. 1 stands fo 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 thi most desirable in children .*

___ *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 c 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 d 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 cate as nominal, ordinal, and numeric .

1.5.1 Data are nominal or categorical when respondents are asked to name, or categorize, their answer. In no 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 Data are ordinal when respondents are asked to rate or order items on a list. In ordinal data, there are r numeric values associated with the categories, and there is no way to measure distance between categori 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 There are two types of numeric data: interval data and ratio data, although statistically these two types tend to be treated the same. With interval data, the distances between the response values (i.e., numbers) real meaning.

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

Measurements of physical properties have characteristics whose quantity or magnitude can be measur ratio scales. Ratio measurements have a true zero, unlike interval data, and comparisons between data pc 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) . The choice of mode v 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 tele and face-to-face interviews) or the respondent enters the data (as in Web surveys and paper-and-pencil su

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 priva interviewer administered modes.

1.7 Review the literature on various techniques that can be used in survey questionnaire design to decrease soci desirable reporting, which can cause social disability bias. For example, random-response technique (RRT) is method designed to elicit reliable responses to sensitive survey items, although it is only useful for a very limi number of yes/no questions in any given survey. In RTT, respondents are randomly assigned to answer one of yes/no questions: one sensitive and the other non-sensitive and with a known probability. The interviewer is u of which question is given to the respondent, and only records the answer. Based on the probability of selectir sensitive question, the probability of respondents who answer yes to the nonsensitive question, and the propor 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 l women, RRT generated the most reliable data regarding induced abortion, when compared to face-to-face sur both audio and paper-based self-administered modes .

1.8 Become familiar with the literature on questionnaire design for comparative surveys and research on the eff culture on how respondents understand and answer survey questions. Respondents' social reality and cultural framework shape their perceptions and survey responses in a variety of ways (see and ). Key areas of this liter include:

1.8.1 General guidance and resources for comparative surveys .

1.8.2 Conceptual challenges, functional equivalence, and comparability .

1.8.3 Operational challenges .

1.8.4 Application of total survey error (TSE) and potential sources of measurement error in 3MC surveys .

1.8.5 Theory integrating culture into models of survey response . For overviews, see and .

1.8.6 Response styles and response bias .

1.8.7 Effect of the research context resulting in order effects ; the presence of third parties during the intervi

1.8.8 Particular questions shown to be vulnerable to cultural effects (e.g., self-rated health ).

1.8.9 Particular types of questions shown to be vulnerable to cultural effects (e.g., subjective probability que

***Lessons learned***

1.1 Cross-cultural questionnaire design literature, while growing, can be hard to locate, unclear, or very sparse details. Even detailed study reports might be clear to people who were involved in a project, but not clear eno 'outside' readers. Detailed and transparent documentation of the questionnaire design process is critical for cro cultural survey research in order for other data users to understand the data collection procedures in each cour to evaluate the data quality in a comparative manner.

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

1.3 discuss multiple dimensions of sensitivity as a broad contextualized concept including social desirability, threatening questions, and privacy concerns. They argue that understanding the reasons for sensitivity and, in particular, whether they are the same or different in different countries can help in the evaluation of potential on data comparability and offer strategies for detecting and addressing differences in question sensitivity in a comparative context.

1.4 Work re-evaluating a series of classic split-ballot questionnaire experiments previously conducted in monoc settings in an online multinational study by indicates that general guidelines developed in the United States (s the question order effect) may apply to other countries. Respondents across 14 countries evaluated the two ab questions differently and the order in which the questions were posed was important despite the high acceptar towards abortion in Nordic countries.

1.5 Research by finds evidence of acquiescence response style (ARS) (the tendency to choose 'agree' or 'yes' re in cross-cultural surveys. Black and Hispanic respondents show more ARS compared to White respondents, v out of three statistical models provide evidence that Black respondents display more extreme response style (I

(the tendency to choose the two endpoints of response scales more frequently than other categories) compared White counterparts.

1.6 Research has shown important cultural differences in responses to subjective probability questions such as ‘ do you think the chances are that [FUTURE EVENT]…?” on a probabilistic scale of 0-100, which are comm used in well-established ongoing surveys in more than 25 countries . Based on cross-cultural psychology liter propose measurement mechanisms and empirically examine the influence of cultural orientations on response several subjective probability questions on different topics in several existing surveys. Overall, draw four mai conclusions from their results: 1) subjective probability questions are difficult to answer for respondents beca involve predictions about future events; 2) apart from cultural background, respondents with Anglo and West backgrounds, among whom the sense of control is typically greater, may experience less difficulty than those lower sense of control (e.g., Spanish-speaking Hispanics in the US), leading to lower nonresponse rates; 3) th sense of control appears to further influence responses, with the lack of sense of control being associated with unrealistic responses such as 0 and 100, resulting in systematic cross-cultural differences in response patterns expressing uncertainly to subjective probability questions may also differ with, in the U.S. context, it being ex through “I don’t know” by minority respondents and 50 by White respondents.

1.7 It is essential to pay close attention to the types of response scales used and the translation of response scale For example, in an analysis of data from five surveys conducted in four different countries (China, Germany, and the United States) asking respondents to rate their health using a balanced scale and an unbalanced scale, that inconsistent translation of the scale point “fair” on the unbalanced self-rated health scales produced differ response distributions, rendering the resultant data less comparable.

↑ Ba

**2. Become familiar with the comparative design options available and the advantages and disadvantages of ea**
***Rationale***

Knowledge of the different approaches available for comparative design for surveys in multiple cultures, languages, 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 advantage disadvantages of the major approaches to questionnaire design for 3MC surveys. The three basic approaches i asking the same questions and translating (ASQT), asking different questions (ADQ) (usually done to adapt t cultural, social, or other needs), or using a mixed approach that combines ASQT and ADQ . 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 c 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 pro 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 suggeste Translation: Overview for more information.
- Decentering is a second way to 'ask the same questions.' With decentering, the same questions are deve simultaneously in two languages—there is no source questionnaire or target language questionnaire. Th decentering process removes culture-specific elements from both versions . However, decentering is on suitable for two-language projects, and its use is restricted in many ways . It is also very work-intensive there is little information about recent experiences using this technique. See for a more detailed explana 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 suita appropriate, or may not be possible. For example, the European Social Survey (ESS) had no issue using to translate "Do you consider yourself as belong to any particular religion or denomination?" into multi languages; however, ASQT was not appropriate for "Do you have difficulty walking several blocks?" U ASQT for the latter was judged problematic in study countries where neighborhoods are not organized blocks.
- Anchoring vignettes allow researchers to make adjustments when respondents from different cultures, c or ethnic groups interpret questions in different ways . Participants are asked to provide assessments bo themselves and for several hypothetical people. Anchoring vignettes assume vignette equivalence (i.e., hypothetical situations portrayed in the vignettes are viewed equivalent across cultures to be compared) reporting consistency (i.e., respondents rate their own and hypothetical vignette persons in a consistent They can be analyzed by both nonparametric methods and model-based parametric methods. However, profile of hypothetical persons in the vignettes can potentially affect the adjustments . Methods have be developed for the evaluation and selection of anchoring vignettes for a diverse range of topics ; see for information and resources. Additionally, recent research on designing anchoring vignettes in a 3MC co stresses the need to control for reporting heterogeneity, and the utility of using anchoring vignettes to d based on survey data from Sweden, the United States, and China.

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, di indicators used in each location are assumed to tap a construct that is shared across populations. For exa following questions may all be effective indicators of the concept of intelligence for individual populatic However, characteristics of intelligence may be more or less salient, depending on local context and AD be the best strategy . 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 tc 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 q on education, for example, are often asked in terms of a shared question stem (such as "What is the hig 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: Shared La Harmonization.
- For cross-cultural surveys, cultural adaptation of instruments along with translation improves measuren comparability . 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 Ap A).

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

2.4 Match the question style (responses) to respondent recall style. For example, incorporate calendar technique event history calendars; see Data Collection: General Considerations) for people who identify time by events 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 requ all impose constraints. For example, if questions from other studies are to be used again ('replicated'), only an model (perhaps with adaptation) is possible for these questions. The chosen data collection method, the sampl design, the fielding schedules, and available funds or stipulations on the use of funds can all limit options (see Design and Organizational Structure, Data Collection: General Considerations, Sample Design, and Tenders, and Contracts).

2.2 Researchers are usually most familiar with the ASQT approach, but may not be aware of the limitations and constraints of this approach . In addition, pressures to use replicated questions might over-promote the ASQT approach. Please see Appendix A for the pros and cons of ASQT.

2.3 Comparability or equivalence is sometimes judged on the basis of similar wording across questionnaires. Tl 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 Eng question “Does he like adventures?” in French is more likely to be understood as “Does he like amorous adve Bilingual or multi-lingual researchers with substantive knowledge of two or more cultures and languages are in this approach. In addition, qualitative study and cognitive testing are critical for questionnaire translations. the mutual understanding among the respondents is the goal.

2.4 It is difficult to find examples of surveys with most substantive questions based on an ADQ approach. Ther examples of research that analyzes different questions from different studies and takes them to reflect aspects given common construct .

2.5 Change of question formats or adapted questions can radically affect respondents’ answers in cross-cultural such as in the International Social Survey Programme (ISSP) (see also Adaptation).

2.6 Researchers also need to be aware of the negative consequences associated with inappropriate standardizati and ).

↑ Ba

**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.** ***Ratio***

Good questionnaires can rarely be developed by a single person. This is especially true for 3MC research. In accord with a quality assurance framework for design, a team is needed that provides the spread of knowledge, diverse skil cultural backgrounds for which successful comparative design calls .

***Procedural steps***

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

3.2 Identify a lead person on the design team who is also responsible for coordinating with other groups involv 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-conferenci including online meetings). Meet regularly to communicate progress.

3.4 Identify the various skills required for the team.

3.4.1 These include all the skills needed for questionnaire design in general, including but not limited to 3M 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 popu structure and needs, etc.).

3.4.3 Depending on their roles on the team, members may need to be conversant in whatever lingua franca i 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 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 .

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 b qualitative and quantitative methods .

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 involve ensures the availability of expertise on given topics and local knowledge. A drafting team might need specific short-term input from an expert on a substantive area in the questionnaire. For example, if input on pensions i needed, an expert on the topic may be brought in exclusively for the development of pension-specific questio

***Lessons learned***

3.1 In addition to the lead team, each cultural group can benefit from strong input from local participants who a similar to the intended sample population. Ways should be found to have any groups who are participating in project, but are not directly part of the core development team, contribute to the development of the questionn This can be a less formal team of local participants who can help guide questionnaire development from the g up, along the lines of "simultaneous [questionnaire] development" . Another option is for the working group t specify target variables while allowing local participants to specify the particular questions . It will be helpful 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 loca community and experiences of the target population, which researchers alone may not be able to recognize or Findings from qualitative methods can be used to inform questionnaire design and subsequent interpretation c quantitative results .

↑ Ba

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

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

While different studies follow different design models (ASQT, ADQ, mixed approaches), this guideline identifies sc 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 translati

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/langu

4.3 Create a schedule and budget for the milestones, deliverables, and procedures involved in the chosen desigr In the ASQT model, this would include schedules and a budget for producing draft source questionnaires, rev participating cultures or groups, deadlines for feedback, schedules for pretesting, schedules for language harmonization, schedules for translation (See Translation: Scheduling), and subsequent assessment and pretes participation of team members from all countries throughout the process is essential in ensuring that the quest 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 ad of outputs (see Survey Quality).

4.5 Create communication channels and encouragements which ensure that participants can and do make feedb 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 leadir figures or countries in either the design team or the larger project. It is important to make clear to participating that every contribution is valuable. Sharing feedback across the project underscores the value of every contrib and explains to participating groups why their suggestions are or are not incorporated in design modifications

↑ Ba

**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 suital a source questionnaire for multiple other versions, rather than as a questionnaire for a single population. Pretesting c 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 for suitability, ideally with every target population in the study. Various qualitative and quantitative approaches can to pretest the target questionnaires in the target population. An often-cited recommendation is “if you do not have th resources to pilot-test your questionnaire, don’t do the study.” .

***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 suitat new context and for use with new populations.

5.2 Whenever possible, pretesting of the source questionnaire should be combined with pretesting a spread of o languages representing the diverse target populations in the project .

5.3 Ensuring the quality of questionnaire development prior to pretesting is just as important as pretesting itself team selection, adequate briefing on requirements and expectations, and good use of documentation will enha quality of the questions presented for pretesting so that pretesting serves the monitoring and refining purposes 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" . In 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 instrument for local suitability .

↑ Ba

**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 cri establishing quality assurance and quality monitoring steps for the process of developing any questionnaire .

***Procedural steps***

6.1 Be cognizant of possible sources of survey error in the questionnaire design phase, and develop quality assu and quality monitoring steps to address these (see Survey Quality). Possible sources of error in this phase incl validity and measurement issues .

6.2 Acquaint question designers with important quality assurance literature on the topic of question design (e.g. validity, tests of conceptual coverage, response process, sources of measurement error) .

6.3 For source questionnaires, form a team in each country or location that meets to discuss the development ar 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 time fashion during the development phases or, as appropriate, report to the coordinating center .

***Lessons learned***

6.1 Quality assurance and quality monitoring should be addressed early in the design planning process. describ explicit procedures and protocols he and his research team followed in designing a multinational study with n existing survey items to measure underlying theoretical concepts of the research question. Because error due t 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 research project. All meetings involved collaborators—experts in a wide variety of relevant disciplines—from study country. Meeting discussions, item wording decisions, and inconsistent field results pointing to measure

issues were carefully documented, and necessary deviations from between-country comparability were detail dataset codebook for future users . See also  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 c design need to be thoroughly investigated and understood when setting quality standards. Some locations or c will need more assistance than others in understanding the relevance of some requirements. They may also ne 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 on these matters.

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

↑ Ba

**7. Develop qualitative and quantitative protocols and procedures for assessing the quality of questions across implementations.** ***Rationale*** Identifying standards to be met and establishing the criteria required to meet them, as agreeing on the good/best practice procedures to follow, are basic to undertaking quality assurance and quality moni

***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 .

7.2 Include qualitative and quantitative methods of assessment (see the Pretesting chapter for a detailed descrip 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 ex translators.
- Debriefings from any testing (interviewers and respondents).

7.2.2 Quantitative methods of assessment include pilot studies :

- Reliability.
- Exploratory and confirmatory analyses such as variance analysis, factor analysis, multitrait-multimetho response theory, latent class analysis, differential item functioning, or standalone experiments or embed 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 . Howev effort has been devoted to comparative experimental study on survey design and translation (for examples, se

7.4 Consider using advance translations or translatability assessments as part of questionnaire design to minimi translation problems.

***Lessons learned***

7.1 Both qualitative and quantitative methods of assessment are necessary. Reliance on one without the other is advised.

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

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

↑ Ba

**8. Develop a documentation scheme for questionnaire design decisions, design implementation, and quality as protocols.** ***Rationale***

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

***Procedural steps***

8.1 Design the documentation process before question development begins, and document question design fron 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 that is user-oriented and user- Include:

8.4.1 Conceptualization from concept to questions.

8.4.2 Operationalization (approach; mode; development across versions; adaptation agreements; annotation (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 qu 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 a 'source version 1.' Documen 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 ad them.

8.2 If documentation is left until 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 questior and translation have been completed) sometimes contain documentation on translation challenges for two or t phrases. The templates used in recent German ISSP translation discussions note a myriad of challenges .

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 quest drafting coordinator or, as appropriate, the coordinating center, the problems can be addressed. For example, f to questionnaire drafting groups from countries participating in the ISSP and ESS studies sometimes leads to in draft versions of source questions.

8.4 At a later stage, documentation might be helpful in understanding potential differences in the data, either ov 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 document

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

↑ Ba

**References**

{2265844:E7G4RK32};{2265844:KIP6EVCR};{2265844:AU6ZLHGK};{2265844:UG9RBUDL};
{2265844:SZHXFRU8};{2265844:5RGK6GYC};{2265844:BLZRKBZS};{2265844:JKYKPQF5};
{2265844:KIP6EVCR};{2265844:Q3VF9MH2},{2265844:5MKBB342},{2265844:8X5QJLMK};{2265844:M8Z
{2265844:YHPXI3LA};{2265844:UG9RBUDL},{2265844:SZHXFRU8},{2265844:KACAY26L},
{2265844:5RGK6GYC},{2265844:BLZRKBZS},{2265844:ZRGVSENQ};{2265844:8X5QJLMK};
{2265844:6LKYYAAW};{2265844:I4VKNHZF};{2265844:E7G4RK32};{2265844:GLTPZNKG};
{2265844:6UEXNEYD};{2265844:RTGSKHIJ},{2265844:KACAY26L},{2265844:W4EUJIW7};{2265844:TU7
{2265844:W5ITP73H};{2265844:FPSCV3U8};{2265844:ND29X4EB},{2265844:UQRNF4QU},{2265844:ZUB
{2265844:8439KHLJ};{2265844:BEZ8Z5KX},{2265844:J3QVK33R},{2265844:QDY9NYEZ},{2265844:C8E6
{2265844:KPBNXP5T},{2265844:KKSJ3KBH},{2265844:86RCWY5J};{2265844:KIP6EVCR},{2265844:KKSJ
{2265844:RCK4HJVV},{2265844:86RCWY5J},{2265844:BWMUUHCE};{2265844:IG9U5DAR},
{2265844:FUCWB8MS},{2265844:QABREUFG};{2265844:BWMUUHCE};{2265844:SBQNWUE9};
{2265844:7DM2CAMH},{2265844:KJTJE2HE},{2265844:IG9U5DAR},{2265844:QABREUFG},
{2265844:TM6KVTXG},{2265844:FPSCV3U8},{2265844:WRAFGSSU};{2265844:QF4QCFUI};
{2265844:EM2JC7TM};{2265844:QF4QCFUI};{2265844:ECM3HGEB};{2265844:5P2WUVHT};
{2265844:K44VUZS5};{2265844:5QIU5QEH};{2265844:WRAFGSSU};{2265844:ECM3HGEB};
{2265844:ECM3HGEB};{2265844:ECM3HGEB};{2265844:5SH7RPZ6};{2265844:4S267T9N},{2265844:KIP6
{2265844:AU6ZLHGK};{2265844:4S267T9N};{2265844:4S267T9N},{2265844:KIP6EVCR},{2265844:V88BZ
{2265844:KIP6EVCR};{2265844:U42U2723};{2265844:N4CTL9RR};{2265844:NIUDHWRC};{2265844:S4GU
{2265844:4H829938};{2265844:4S267T9N};{2265844:YLV2AF3H},{2265844:XF7DEBIT};{2265844:JCAXLU
{2265844:7YQDECHB},{2265844:KIP6EVCR},{2265844:AU6ZLHGK},{2265844:6X7KLZ7U};{2265844:6G7
{2265844:G7G2QBMA};{2265844:N27PWFSB};{2265844:UZMQ3EN4};{2265844:DUHNMIAC};
{2265844:PASYR9YW};{2265844:ZX3V7DLE};{2265844:KIP6EVCR},{2265844:AU6ZLHGK};
{2265844:6LKYYAAW};{2265844:I6WNVJ5P};{2265844:SHBPSS9Y};{2265844:78JN8ZY3};{2265844:W2CS
{2265844:CUBTBHA2};{2265844:HNAKUCIJ};{2265844:DDYLNLET};{2265844:DDYLNLET},
{2265844:M8ZJBZXV};{2265844:DUHNMIAC};{2265844:NA6XYIV9};{2265844:NA6XYIV9};
{2265844:GT3JYQM4};{2265844:DDYLNLET};{2265844:M8ZJBZXV},{2265844:BGCZP7E8};
{2265844:EF39Q2YA},{2265844:TRA6TD6N};{2265844:RQS4BYQJ};{2265844:7YQDECHB} apa creator asc

↑ Ba