

## Introduction

### Guidelines

1. Identify what the pretest should achieve and choose a pretest design that best fits the study goals and ea population.
2. Combine pretesting techniques to create a comprehensive design plan that takes advantage of the streng minimizes the weaknesses of each method.
3. Train or hire staff members who are able to adequately implement the chosen pretesting technique(s).
4. Conduct the pretest in the same mode of data collection (interviewer administered or self-administered) main survey.
5. Conduct the pretest with the same target population as the target population for the survey.
6. Evaluate the results of the pretest.
7. Fully document the pretesting protocol and findings.

### References

### Further Reading

---

## **Introduction**

Pretesting plays an essential role in identifying, and potentially reducing, measurement error that damages statistical estimates at the population level and thus endangers comparability across populations in multinational, multiregional multicultural surveys (which we refer to as '3MC' surveys). Pretesting involves a variety of activities designed to evaluate survey instrument's capacity to collect the desired data, the capabilities of the selected mode of data collection, and overall adequacy of the field procedures. Throughout this text, we refer to a 'pretest' as the collection of the qualitative and quantitative techniques and activities that allow researchers to evaluate survey questions and procedures before data collection begins. [Table 1](#) provides a summary of the most commonly used pretesting techniques, such as pilot studies, cognitive interviewing employing concurrent or retrospective think-aloud techniques, focus groups, behavior coding, and more. For more on cognitive interviewing, please see the [Cognitive Interviewing](#) subchapter.

As suggested in the survey lifecycle, many pretesting activities take place once the questionnaire and other survey materials have been developed, adapted, and translated. However, pretesting techniques such as focus groups and vignettes are often used in advance of the overall research and questionnaire design in order to inform question wording and other aspects of the research design (appropriate target population, data collection mode and procedures, etc.).

'Pilot studies,' also referred to as 'dress rehearsals' or 'field tests,' encompass pretesting procedures that employ all the procedures and materials involved in data collection (regardless of how small of a scale) before the actual data collection begins. They are typically used to achieve one or multiple specific goal(s)—from estimating response rates under a particular recruitment protocol to identifying an optimal design characteristic (e.g., incentive amount) through experimentation. Below, we identify the following as reasons for conducting a pilot study:

- Checking the length of the instrument or interview relative to the culture of interest.
- Checking adaptations of instruments.
- Checking the target population's familiarity with units of measure (e.g., currency, Imperial vs. metric system, etc.).
- Checking the target population's familiarity with constructs and concepts (e.g., proper names, 'hamburgers,' etc.).
- Checking the target population's familiarity with the instrument layout.
- Identifying the customary answering process in the culture of interest (e.g., checking boxes, circling answers, etc.).
- Comparing item difficulty statistics (for example, see [Q-Bank](#) and [SQP](#)).

Researchers often draw on a combination of qualitative and quantitative methods to test draft questionnaires and other materials. Using qualitative methods for an overall mixed-methods instrument design serves as a process of integration (often iterative) design and pretesting.

This chapter provides examples mainly based on U.S. surveys that sample ethnic minorities and immigrants and are administered in different languages, but attempts to extrapolate experiences and lessons learned to cross-national surveys.

When multiple languages are used in the same survey, pretesting the different language versions is an essential part ensuring measurement equivalence and cultural and cross-cultural equivalence (see [Translation: Overview](#)). In addition, it is often difficult to employ the same mode of data collection across countries participating in a cross-national project. It is important to test in advance the suitability of the selected mode for the survey topic and population (see [Study Design: Organizational Structure](#)). Pretesting techniques may have limited application in a given context and culture. Research on how pretesting strategies may need to be tailored to suit different populations is only beginning to be undertaken systematically. See [Pretesting](#) for a discussion.

[↑ Back to top](#)

## Guidelines

**Goal:** To ensure that all versions of the survey instrument adequately convey the intended research questions and measure the intended attitudes, values, and reported facts and behaviors, and that the collection of data is conducted according to specified study protocols in every country and in every language.

[↑ Back to top](#)

### 1. Identify what the pretest should achieve and choose a pretest design that best fits the study goals and each population. *Rationale*

Determining what issues have to be addressed allows for the best use of the various pretesting techniques—whether researchers want to test all field procedures, or only the survey instrument (or parts of it) or the equivalence of the survey instrument across languages and modes of data collection. Pretesting for a study may combine several complementary pretesting techniques (see below) and should be done in each country participating in the research. Even if some or all questions have been used in other studies, pretesting for the local context is necessary to assess their performance in the mode and question order of the current study and with the target population and the performance of the translation.

[Table 1](#) summarizes the most commonly used pretesting techniques with a brief description, a list of their strengths and weaknesses, and the context in which each is typically used.

#### *Procedural steps*

- 1.1 Using [Table 1](#) as an aid, decide what pretesting technique(s) will best fit the study's purpose.
- 1.2 Consider the cultures within which the study will be conducted and, where possible, establish standardized pretesting protocols across countries regarding:
  - 1.2.1 How to best convey the objective of the task.
  - 1.2.2 How to standardize or harmonize the pretesting protocol.
  - 1.2.3 How to select staff members for the pretest.
  - 1.2.4 How to train staff.
  - 1.2.5 How to monitor quality. Audio and video recordings are often made during cognitive interviews and focus groups to help with the reporting process, and such recordings can also be used to monitor interviewers and group moderators to ensure adherence to the pretesting protocol guide. Computer-Assisted Recorded Interviewing (CARI) allows for monitoring during field pretesting and field data collection to detect interviewing errors.

fraud and ensure data quality . For a larger discussion on the importance of quality control and how to incorporate it at various survey stages, see [Survey Quality](#) and [Paradata and Other Auxiliary Data](#).

- 1.2.6 How to analyze results of the pretest (e.g., whether the analysis will be qualitative and/or quantitative)
- 1.2.7 How to report and address problems.
- 1.2.8 How to decide on changes to the survey instrument.
- 1.3 After selecting a pretesting technique:
  - 1.3.1 Assess whether to conduct the pretest(s) in-house or to contract the testing to an outside organization.
  - 1.3.2 Establish a time schedule that adequately matches the pretesting design, allowing sufficient time to implement any revisions which may be deemed necessary based on results from the pretest prior to implementing the study.
  - 1.3.3 Budget accordingly. Be sure to include expenses related to interviewer and staff training, respondent recruitment, and incentives, if applicable, for pretest subjects.
  - 1.3.4 Plan how to document the procedures and findings and how to best share them with teams in other countries.

### ***Lessons learned***

- 1.1 In 2012, the German Data Forum established an expert group to provide minimal requirements for assessing and documenting the measurement quality of established and newly developed survey instruments. Six quality standards were derived for each stage of the measurement process. presented these quality standards for survey instruments contrasted them with existing alternative standards from other countries and/or disciplines .
- 1.2 Available pretesting techniques may vary across countries depending on testing traditions, resources, trained interviewers and respondents' familiarity and experience with them. Even when the same pretesting technique is used, if its implementation varies drastically across countries, it becomes impossible to determine whether observed differences are due to differences in the response process, translation, or the conceptual spectrum. For example, it is not safe to assume that procedures for conducting cognitive interviews will be the same across all countries. Differences exist in the experience of the interviewers, the location of the interviewing, the methods used to recruit participants, the approaches to creating the interviewing protocol, and respondents' experience with cognitive interviews. I work in seven countries and eight languages has focused on creating a common approach to cognitive interview questions designed to measure health status . To ensure equivalence, all parties involved in the project agreed on a method to be used for recruiting participants, administering the protocol, and documenting results.
- 1.3 Even when standardized protocols are used across countries, pretesting techniques such as cognitive interviews do not always work equally well across cultural groups without modification . investigated the efficacy of concurrent think-aloud as a pretesting strategy with Chinese respondents. Her investigation identified challenges and limitations of taking methods developed in one language and culture and directly applying them to another. She points to the need to include consideration of sociolinguistic conventions appropriate to different cultural groups when conducting cognitive interviews, because cognitive processes in survey interviews are influenced by cultural background encompassing language. Some recent studies have examined ways of improving the cognitive interviewing experience for Spanish-speaking respondents in the United States and respondents outside of the United States and Europe. Furthermore, stress the relationship of rapport-building and how the cognitive interview is introduced to Chinese, Korean-, and Spanish-speaking participants in the United States based on exploratory studies. By helping the interview participants feel at ease and more familiar with the cognitive interviewing task and the purpose of the interview, they seem to better understand the probing questions and feel more comfortable describing their thoughts and opinions.

- 1.4 Culture- or language-specific probes may be needed to test the translation and/or adaptation of a survey instrument. The Census Bureau conducted cognitive tests of the translations of introductory letters and informational brochures for the American Community Survey in seven languages. The focus of the study was to examine how cognitive interviews work in non-English languages given cultural differences in communication. Remarkable differences in the way participants from different language groups provided responses were reported. Chinese and Korean respondents tended to provide limited responses and their answers were not focused on the topic; Russian respondents showed a tendency to always give 'confident' answers; Spanish and Chinese respondents tended to repeat questions verbatim when asked to paraphrase them. Such differences in response patterns raise questions related to data quality and the comparability of cognitive interview results across language groups.
- 1.5 In surveys where monolingual and bilingual respondents are targeted as primary users of a survey instrument, it is recommended including both monolinguals and bilinguals in survey translation pretesting. Their findings show that bilingual respondents identified most of the same problems as monolingual respondents, but not necessarily with the same frequency, and that a small number of issues were only problematic for either the monolingual or bilingual respondents.
- 1.6 In addition to standard pretesting methods, which focus on question wording and format, ethnographic pretesting techniques may be used to identify shared cultural characteristics. Ethnographic techniques emphasize cultural variables, such as belief systems and everyday practices, which determine whether or not a question makes sense within the culture.
- 1.6.1 Consensus panels are similar to focus groups, but are more structured and limit discussion among participants. A panel of people is selected for their expertise and other characteristics deemed to be relevant, and they are invited to answer one or more questions about which there may be considerable doubt or disagreement in order to see if a consensual view can be reached.
- 1.6.2 Questerviews are standardized self-completed questionnaires administered within the context of an in-depth qualitative interview. Respondents are encouraged to discuss their definitions of terms and responses to items while they complete the standardized questionnaire. Usually, questerviews are tape-recorded and transcribed for analysis to identify emergent themes.
- 1.6.3 Ethnographic pretest interviews ask broader questions than cognitive interviews and may be used to identify additional terms regarding a domain of interest and to identify cultural schemas. They are unstructured, nondirective interviews that focus on understanding the interviewed individual's cultural background so that questions are appropriate to that individual's life. Gerber recommends asking ethnographic questions after completing the regular cognitive interview. offers the following examples of probes which may be used with various cultural groups:
- "Tell me about the types of activities you do that take physical effort or that make you feel physically tired."
  - "The question has a list of foods in it. Are these the types of foods that your family usually eats?"
  - "What types of things do you think of as 'work'?"
  - "Are you always paid in cash for the work you do, or are there other ways in which you get paid?"
- 1.7 A related practical question is whether to create cognitive protocols in English and then translate into the target languages, or to develop the protocols directly into the target languages, accounting for different cultural norms and socialization styles. Each approach has benefits and weaknesses that must be weighed against one another given specific survey conditions (e.g., simultaneous development of the protocol guides may not be as feasible in multilingual projects as it is in bilingual studies). This section discusses approaches for creating protocol guides in multiple languages that ensure culture and language appropriateness and present strategies for respondent recruitment, interviewer selection and training that allow adequate testing of instrument translation.

- 1.8 In the pretesting stage, in addition to identifying issues in the questionnaire, cognitive interviewing can also be used as a study of construct validity in comparative surveys. See [Cognitive Interviewing](#) for a more in-depth discussion of this pretesting technique.
- 1.9 While focus groups are a quick way to gain in-depth insight into participant knowledge and attitudes, argue studies, particularly in the health field, are relying too heavily on this technique. While previous research has shown that focus groups are generally useful in collecting information of a sensitive nature, some topics are exceptions. In a case study in Malawi, adolescent girls were interviewed using two different methods: in-depth interviews and group discussions. The study, conducted through the National AIDS Control Programme, utilized mixed methods through quantitative data collection of census information and highly-structured questionnaires as well as qualitative observation, less-structured interviews, and focus groups. Overall, the study found that studies cannot solely rely on focus groups because some topics are so sensitive that individuals will not discuss them in front of one another. For the female subjects in Malawi, menstruation was too sensitive to discuss in focus groups. The authors recommend that researchers use both methods, with in-depth interviews conducted before focus groups. They found that by asking females sensitive questions during their in-depth interviews, they were then able to follow up some of the interview questions by asking if the subject would be willing to discuss this topic in groups of girls.

[↑ Back](#)

## **2. Combine pretesting techniques to create a comprehensive design plan that takes advantage of the strengths and minimizes the weaknesses of each method. *Rationale***

Pretesting techniques often complement one another and can logically be combined to maximize the efficiency of the design (see [Table 1](#)). For example, to minimize cost, one can consider pretesting a questionnaire using expert review; the questionnaire is revised based on reviewers' comments, participants for cognitive interviews can be recruited or the study can be launched. On the other hand, studies comparing multiple pretesting methods have found that different pretesting methods produced different and sometimes even contradictory results regarding the performance of survey questions. Therefore, it is of great importance that techniques are selected with sufficient consideration of each candidate method's strengths and weaknesses.

In addition, it is important to take language, cultural norms and traditions, and interviewer characteristics (see [Data Collection: General Considerations](#) and [Interviewer Recruitment, Selection, and Training](#)), into account when choosing pretesting methods. The most appropriate combinations of pretesting techniques may vary across countries involved in the study. This should be taken into account when results from the different pretests are evaluated and compared.

### ***Procedural steps***

- 2.1 Begin with pretesting methods that focus on specific aspects of the study (for example, wording of particular questionnaire items, comprehensibility of the informed consent, procedures for interviewers to follow in administering the survey) before moving on to techniques that pull all aspects of the project into a more comprehensive study.
  - 2.1.1 For example, consider a focus group or in-depth interviews for initial development of constructs, cognitive interviews for questionnaire development and refinement, and a field pilot study for an overall test of the instrument and field procedures. Often, a pilot study with a robust sample can be the best way to test the instrument, and data analyses with sufficient power can be the most effective way to ascertain if the questionnaire is working as intended.
- 2.2 Discuss every round of changes introduced to the questionnaire with the coordinating center and test again—consider several iterations of testing, rather than one large scale pretest.
- 2.3 Be prepared to do multiple rounds of pretesting.

### ***Lessons learned***

- 2.1 In preparation for the shift from a paper-and-pencil instrument to a computer-assisted instrument incorporating a large audio computer-assisted self-interview (A-CASI) component, the U.S. Substance Abuse and Mental Health Services Administration (SAMHSA) implemented a comprehensive pretesting plan. The overarching goal of pretesting was to develop an optimal computerized instrument on the sensitive topic of drug usage. It was also essential that any differences in reporting due to the mode change to A-CASI be identified so that data users would understand how to interpret trend lines from the data. Pretesting work first concentrated on small-scale cognitive laboratory testing to determine the best way to structure the instrument, to train respondents to use the computer, the A-CASI components, to determine the voice to be used for the audio component, and to assess respondent willingness to enter different types of data into the computer (e.g., open-ended responses). Based on results from these laboratory studies, a pilot study was conducted to evaluate interviewer training materials and to collect sufficient data to determine how the mode change impacted reporting. After changes were made based on this field pilot study, a pilot study incorporating an experimental design was conducted. Finally, the revised instrument and procedures were implemented in a split-sample comparison with the original paper-and-pencil instrument during data collection to allow researchers to assess the impact on the trend lines.
- 2.2 The General Social Survey (GSS) does a 'full pretest,' which tests all new items in a realistic field situation with representative respondents, between cognitive pretesting and a pilot study.

[↑ Back](#)

### **3. Train or hire staff members who are able to adequately implement the chosen pretesting technique(s). *Rat***

The selected pretesting procedures may require skills not possessed by the available interviewers. For example, cognitive interviewing requires a discursive interviewing style which is different from traditional standardized interviewing and requires additional training. Sufficient time and effort should be allowed to train staff members and develop protocols that correspond to the selected pretest design.

#### ***Procedural steps***

- 3.1 Select staff members who are fluent in the language of the pretest and sensitive to cultural and linguistic nuances. When different pretest designs are employed in different countries, select interviewers, training, and protocol that match the chosen technique; when the same techniques are used in various countries, harmonize all procedures.
- 3.2 Train staff members for the pretest.
- 3.3 Consider interviewer characteristics as they may affect the outcome of a pretest in some cultures more than others (e.g., conversational styles in many cultures are largely determined by the education, gender, or status of the interviewer and the social hierarchy).
- 3.4 Monitor interviewer behavior to ensure data quality.

#### ***Lessons learned***

- 3.1 Ample time is needed to train local interviewers who may have little or no experience with cognitive interviewing. In the World Health Organization Model Disability Survey, five half-days of training were scheduled to train 10 Nepali interviewers on how to conduct cognitive interviews. However, early on in the training, it became apparent that even though the interviewers were experienced in standardized interviewing, cognitive interviewing was a new concept. The interviewers had difficulty shifting from standardized interviewing to the protocol of probing the respondent for think-aloud answers. A training day was added to the agenda to give the interviewers extra practice with the probing protocol. The interviewers also had difficulty understanding that getting the respondent to give a response was less important than knowing what the respondent was thinking when formulating their answer. This became apparent after several cognitive interviews were completed. During a daily debriefing, an interviewer



that a respondent was having difficulty giving a codable answer, and that she probed until she received one but to probe what the respondent was thinking.

[↑ Back](#)

#### **4. Conduct the pretest in the same mode of data collection (interviewer administered or self-administered) as main survey. *Rationale***

Whatever the eventual mode of data collection, the early stages of research design—testing the construct itself—typically uses face-to-face laboratory methods such as focus groups, cognitive interviews, or vignettes. (See [for a discussion of developing an instrument prior to testing that instrument.](#))

Once a draft questionnaire has been developed, however, it should be tested in the same mode of data collection as the survey. There are several significant differences between interviewer- and self-administered surveys. Respondents listen to the questions in interviewer-administered surveys, whereas they read the questions in self-administered surveys. Interviewer-administered surveys involve social interaction between the interviewer and the respondent; self-administered surveys do not. In interviewer-administered surveys, the interviewer handles routing through the questionnaire; self-administered surveys require the respondent to navigate through the questionnaire. Interviewer-administered and self-administered questionnaires also produce different context effects (e.g., recency and primacy) and may also result in differences in socially desirable responding (see [Study Design and Organizational Structure](#) and [Data Collection: Face Surveys](#)). In order to determine how well proposed procedures will work in the field, pretesting should be conducted in the same mode as the final survey.

##### ***Procedural steps***

- 4.1 If different modes of data collection are going to be employed across countries, pretest in the respective mode.
- 4.2 Some pretest techniques are not portable across modes (for example, behavior coding); others require modification. Adapt pretesting techniques to better match the mode of survey data collection (e.g., ).
- 4.3 Use the latest version of the instrument and the respective materials (e.g., show cards, event history calendar).
  - 4.3.1 Use version control to manage revisions to documents and other materials.
- 4.4 Use field administration procedures planned for production data collection.

##### ***Lessons learned***

- 4.1 Since each mode of data collection has its specific characteristics, it is important to pretest the survey instrument and procedures in every mode that will be used, whether or not the survey questionnaire is translated to a different language. In fact, a change in mode may necessitate changes in wording or changes in design in order to achieve measurement equivalence. For example, cognitive testing for the 2001 U.S. Census showed that more redundancy was needed in the instructions for the “respondent race” question in order for the respondents to be able to follow the “select one-or-more” option in telephone administration. A slightly reworded version of the instructions and system resulted in better understanding of the intent of the question over the phone compared to what was needed when asking the question as it appeared in the mail questionnaire.

[↑ Back](#)

#### **5. Conduct the pretest with the same target population as the target population for the survey. *Rationale***

To most effectively pretest the survey instrument or field procedures, pretest respondents from the intended target population, or, if appropriate, a sub-group within the target population. Ideally, the natural flow of the survey instrument should

tested for each culture and language to avoid awkward conversational situations, question order with unpredictable (or) dependent context effects, question repetition not intended in the source, or other culture-specific problems. The population of a pilot study should be an adequate reflection of the survey target population. For example, if the survey design involves oversampling of certain ethnic groups, the pretest sample should also include reasonable representation of these groups. A pretest with sample persons from the target population will most accurately reflect what will happen during actual data collection in terms of cooperation, respondent performance, total interview length, questionnaire performance, survey cost, etc.

### ***Procedural steps***

*For all pretesting techniques:*

- 5.1 Tailor subject or respondent recruitment to the population of interest.
- 5.2 Prepare all necessary materials that would be used in the main survey, including an informed consent form that reflects the goals and risks of the pretest study (which may be different from the main survey).
- 5.3 Select a sample size that is suitable for the chosen pretesting method.
- 5.4 Apply quotas or use a random sample of the target population to control the demographic make-up of the sample.
- 5.5 Monitor pretest participant recruitment to ensure best use of the chosen pretesting method.

*For pilot studies:*

- 5.6 Select a sample large enough to provide sufficient statistical power to answer the research questions identified in your pilot study analysis plan. Allow for nonresponse, noneligibility, etc.
- 5.7 Follow the sample selection protocol planned for the final study.
- 5.8 Monitor the sample selection.

### ***Lessons learned***

- 5.1 Select respondents from the survey target population; however, keep in mind that sometimes 'survey-trained' respondents may be needed to detect potential problems. A study on pretesting by Dillman (1978) demonstrated that the general population may not be a good judge of the quality of survey questions, even when this is the target population. Researchers introduced obvious errors in the short questionnaire (e.g., missing response alternatives, inappropriate vocabulary) and asked respondents to be critical of the questions while answering them. Only a third of the sample noticed a missing response alternative; almost no one commented on 'double-barreled' questions or 'loaded' wording. One possible explanation is that all of the respondents had roughly the same low level of survey experience.
- 5.2 Work conducted by the U.S. Census Bureau to develop a bilingual (English/Spanish) decennial census form involved cognitive testing to identify potential problems with the layout of the form, to test respondents' ability to correctly navigate through the form, and to assess the quality of the Spanish translation. Testing did not directly assess the English questions, as the wording of the English items had already been nearly finalized. As part of a particular study, cognitive interviews were conducted with monolingual Spanish speakers and bilingual Spanish dominant speakers to focus on translation issues. Results from the testing indicated specific questions that were problematic for Spanish speakers. However, because there was no comparable group of English speakers included in the testing, it was difficult to determine whether the problems were confined to the translated items or would also be problematic for respondents who read the English wordings. To eliminate this problem, in a second round of testing, monolingual English respondents were included as well. The inclusion of these respondents allowed the researchers to identify where problems with the Spanish translation was due to specific choices made in the translation and not to the English questions.



concepts were unclear for the Hispanic respondents as opposed to questions that were equally unclear for both and Spanish speakers.

### 5.3 Large established cross-cultural studies vary in the type and amount of pretesting they do.

5.3.1 Prior to the start of Round 1, the European Social Survey (ESS) source questionnaire was pretested using 'interaction analysis' (i.e., behavior coding) to identify questions which were problematic for the interview respondent. Problem questions were modified, and the questionnaire was translated into various languages in accordance with ESS Round 5 specifications, each participating country was required to pretest its translated questionnaire on a quota controlled, demographically balanced sample of around 50 people. The aims of pretesting were, at a minimum, to check routing and comprehension. Ideally, the pretests could also be used to check for equivalence between the translated version of the questionnaire and the source. Countries were encouraged to audio-record interviews, conduct respondent and/or interviewer debriefings, and use cognitive interviewing to test for equivalence. The specifications note that these pretests occurred after the source questionnaire had been finalized, and that opportunities to amend the source questionnaire were extremely limited at this point.

5.3.2 The Survey of Health, Ageing and Retirement in Europe (SHARE) utilized a four-stage questionnaire development process. In the first stage, working groups produced an English-language draft questionnaire drawn from preexisting survey instruments. The draft questionnaire was piloted in the U.K. in September. Based on the lessons from this pilot, the English-language questionnaire was revised and translated into SHARE languages. In the second stage, the translated questionnaires were simultaneously piloted in all 14 countries, each testing a quota sample of 75 persons. In the third stage, after further revisions to the survey instrument, the full questionnaire was tested in all countries using probability samples (some 100 primary respondents per country plus their spouses). This all-country pretest also tested the country-specific logistics and the procedures to achieve probability samples. During the fourth stage, pilot and pretest results were statistically analyzed, leading to the final design of the questionnaire.

[↑ Back](#)

## 6. Evaluate the results of the pretest. *Rationale*

The goal of the pretest is to identify problems in the questionnaire and study design in each country. The results of the pretest have to be evaluated to determine the best way to fix existing problems without introducing new ones. Changing the survey instrument and design should be considered in the context of the whole study—changes that fix a problem in one country may introduce a problem in another. The coordinating center should decide whether minor differences that preserve the measurement equivalence of the survey instrument across countries can be tolerated (see [Translation: C](#) and [Study Design and Organizational Structure](#)). Any introduced changes in instrument design should also be pretested to avoid unforeseen errors (also see [Instrument Technical Design](#)).

### *Procedural steps*

#### 6.1 Examine the findings of each pretesting technique used, and identify the causes of the any problems discovered.

6.1.1 Decide in advance what constitutes a problem. For example, the 10%+ rule is often used in behavior coding for flag questions: if a question is misread or misunderstood by over 10% of respondents, then it is considered problematic. The appropriate threshold for any particular study is often determined from the distribution of coded errors (which is dependent on the coding scheme and instructions for code assignments).

6.1.2 Look for problems that are common across interviews, but also be aware that a problem may be important even if it occurred in only one interview. This is especially important when qualitative techniques are used. In order to determine what constitutes a problem, all possible factors that play a role in the pretest should be considered.

- 6.1.3 Examine in what situations and with what types of respondents problems occur.
- 6.2 If a pilot study has been conducted:
  - 6.2.1 Review response distributions and item nonresponse for key study variables.
  - 6.2.2 Review interview length.
  - 6.2.3 Review satisficing behaviors.
  - 6.2.4 For attitudinal and value variables, check whether items group together as intended in the survey (e.g., confirmatory factor analysis, latent class analysis (LCA) , analysis of variance ).
  - 6.2.5 Solicit and review feedback from interviewers and respondents.
- 6.3 Report the results and proposed changes to the coordinating center. It is important that the timing and documentation of the pretest are coordinated across participating countries to allow overall comparison of responses and propose meaningful changes.
- 6.4 If changes are introduced to the questionnaire or design procedures, plan for another pretest.

### ***Lessons learned***

- 6.1 Pretesting techniques, and the results they yield, are meaningful only when the selected procedures are culturally appropriate. Few pretesting techniques have been tested and studied across countries; thus, some may not be successfully implemented, leading to meaningless results in certain cultures.
  - 6.1.1 Studies in psycholinguistics, for example, have demonstrated different cognitive tendencies between Chinese and English speakers in counterfactual reasoning . When asked what their thoughts would have been on hypothetical legislation by their government, Hong Kong respondents consistently responded that the government has not proposed such legislation. Chinese speakers were less attuned to hypothetical thinking because their language does not mark counterfactuals differently from conditional statements. Such examples suggest that certain cognitive laboratory methods (for example, vignettes) may be of limited use in some cultures. On the other hand, the use of vignettes suggests that vignettes may help assess the “cultural sensitivity of a questionnaire”.
  - 6.1.2 There are certain error sources that are unique to cross-national questionnaires, or occur less frequently in single-nation studies. Tools that help to identify these errors and separate them from measurement errors only occur in single-nation studies assist the cross-national survey researcher in producing a higher quality source questionnaire. In turn, this supports translators in producing functionally equivalent translations that work well in the target languages and cultures. The Cross-National Error Source Typology (CNEST) was developed as a tool for improving the effectiveness of cross-national questionnaire design, and has proved useful when used to categorizing and analyzing the results of cognitive interviews .
- 6.2 The analysis of some pretesting methods can be very labor intensive. For example, transcription is often required for focus groups and cognitive interviews. Analyzing this type of qualitative data requires extensive effort. One common approach is to review all interviews, looking for patterns, and then randomly select a few cases for deeper analysis.

[↑ Back to top](#)

## **7. Fully document the pretesting protocol and findings. *Rationale***

Providing a permanent record of problems encountered during the pretest(s) and any changes made to the questionnaire, respondent materials, and field procedures aids staff and researchers working on similar studies or on later rounds of data collection.

same study.

**Procedural steps**

*In a manner consistent across countries, document:*

- 7.1 The pretest sample selection and recruitment method, including the sampling frame and sample size.
- 7.2 The use of incentives.
- 7.3 The geographical location of the pretest.
- 7.4 Respondent characteristics.
- 7.5 Mode(s) of pretest administration.
- 7.6 Dates of data collection and organization(s) conducting the interviews.
- 7.7 Types of staff conducting the pretest (e.g., experienced interviewers, supervisors) and the training they receive.
- 7.8 All materials used in the pretest.
- 7.9 Pretest findings and their implications.
- 7.10 Any changes made to the survey instrument, and the pretesting source that lead to these changes.
- 7.11 The number and types of pretests.

**Lessons learned**

- 7.1 The documentation can serve as a resource for future studies. For example, researchers within a U.S. Federal Interagency Group have developed [Q-BANK](#), a database of questions for national health surveys maintained by the Questionnaire Design Research Laboratory (QDRL) at the National Center for Health Statistics, Center for Disease Control (CDC). The database catalogues tested questions and links each question to cognitive testing findings. Questions are searchable not only by content or subject matter (e.g., asthma questions, cancer questions, demographics, etc.), but also by question type (e.g., objective characteristics, behavioral reports, attitudes, etc.) and response category type (e.g., yes/no, open-ended, quantity, etc.), and response error type (e.g., problems with recall problems, etc.). A statistical tool has been developed that performs basic statistical procedures on questions in the database.
  - Q-BANK, when completed, will centralize cognitive testing reports with links to specific questions and topics and will advance the field by: 1) serving as a resource in the development of new questions, 2) allowing question and response error comparisons across studies, 3) performing analysis on the characteristics of questions contributing to specific response errors, and 4) serving as a research tool investigating response error.
  - Q-BANK is available to any interested researcher. Researchers are also encouraged to contribute their own reports to the catalogue to strengthen the utility of the site.

[↑ Back](#)

*Table 1: Pretesting methods and their strengths and weaknesses (these can be iterative and can be used in combination)*

Approach	Pretesting Method	What it is	Strengths	Weaknesses	Most Common
----------	-------------------	------------	-----------	------------	-------------

					Use
Field Methods	Field pilot study (for an overview, see )	A miniature version of the main data collection	Realistic; allows for testing all field procedures; allows for feedback from interviewers, field managers, respondents, and data analysts	Costly; requires large sample size relative to the other techniques; needs to be planned and conducted in advance to allow time for changes	Field work test
	Interviewer debriefings (for an overview, see )	Small group discussion with interviewers to talk about their experiences	Uses interviewers' expertise on what makes a question difficult in a particular situation and with particular types of respondents	Interviewers themselves may be responsible for the respondents' confusion/problem with a question	Field work test
	Respondent debriefings	Respondents' comments on specific questions or the survey as a whole (usually collected during a field pilot study as a separate interview)	Cheap—conducted as part of the field pilot study; allows for identification of question-specific problems; large sample size allows for confidence in results; realistic (field setting)	In some cultures, respondents may not want to admit confusion and inability to understand a question; increases respondent burden as the length of the interview increases; may be hard to recall items that were problematic	Field work test
	Behavior coding (e.g., ; also, )	Systematic coding of the interviewer-respondent interaction in order to identify problems that arise during the question-answer process	Direct observation of the question-answer process; comparability when standard codes are employed; replicable; allows for use of universal codes, but also study-specific; quantitative; requires medium sample size (30 interviews are considered sufficient to detect problems)	Time and labor intensive; requires well-trained coders and consistent use of the coding scheme; does not identify the exact problem in a question with many codes	Questionnaire testing; field management
	Focus groups (see for an overview; also )	Small group of people brought together to discuss specific topics in a relatively unstructured manner, led by a moderator who ensures the flow of the conversation is	Useful when there is no information on the topic of interest; uses the same types of respondents who are the target population for the survey; allows for immediate followup; requires small group size (10–12 participants)	Mainly qualitative; results should be carefully interpreted due to small sample size; requires well-trained moderators; small group dynamics may influence the results	Questionnaire development

		in the intended direction		
--	--	---------------------------	--	--

[↑ Back](#)

Approach	Pretesting Method	What it is	Strengths	Weaknesses	Most Common Use
Cognitive Laboratory Methods (for an overview, see )	Vignettes (e.g., )	Brief stories/scenarios describing hypothetical situations or persons and their behaviors to which respondents are asked to react in order to allow the researcher to explore contextual influences on respondent's response formation processes	Allows for quantitative analyses; suitable for sensitive topics; requires small sample size relative to the other techniques	Disconnect between a hypothetical situation and respondent's actual views and behaviors; cultures may differ in their ability to think hypothetically (e.g., )	Question development; concept understanding
	Concurrent think-aloud (see and )	Respondents' reports of the thoughts they are having while answering a survey question	Open format with potential for unanticipated information; lack of interviewer bias when probes are not used	Unnatural; high respondent burden; may affect the natural response formation process, thus providing an unrealistic picture of how respondents answer questions in the field; coding may be burdensome; assumes respondents are able to identify and report what information they used to come up with a response to the survey question; respondents may begin to overinterpret the questions and come up with problems that do not exist in the natural context	Question development
	Retrospective think-aloud (see )	Interview with respondents after they have completed a	Does not interfere with the response formation process	Assumes respondents are able to identify and report what information they used to come up with a	Question development

		survey about how they came up with answers to specific questions		response to the survey question; assumes information is still available in short-term memory	
--	--	--	--	--	--

[↑ Back](#)

Approach	Pretesting Method	What it is	Strengths	Weaknesses	Most Common Use
Other	Expert review (for an overview, see )	Review of draft materials by experienced methodologists, analysts, and translators	Cost-efficient; quick; can identify a wide variety of problems in the survey questionnaire (from typos to skip patterns); requires very small sample of experts (usually 2–3)	Subjective; no 'real' respondents involved	Question development
	Question Appraisal System (for example, )	A systematic appraisal of survey questions that allows the user to identify potential problems in the wording or structure of the questions that may lead to difficulties in question administration, miscommunication, or other failings	Cost-efficient; provides sense of reliability due to standardization	Identifies a problem without pointing out to a solution	Question development
	Usability Testing (see and )	Testing of the functionalities of CAPI, CATI, sample management systems or printed materials such as respondent and interviewer booklet, show cards, etc.	Direct user assessment of the tools that will be used during data collection; can be inexpensive and conducted with employees of the survey organization; usually requires small sample sizes	Time consuming	Field work test

[↑ Back](#)

Approach	Pretesting Method	What it is	Strengths	Weaknesses	Most Common Use

Statistical Modeling	Multi-trait multi-method (MTMM) Database (see )	Database of MTMM studies that provides estimates of reliability and validity for over 1000 questionnaire items	Provides quantitative measures of question quality	Costly and labor intensive; questions are considered in isolation, so question order effects might be ignored	Questionnaire development
	Item Response Theory (IRT) Approach (see )	Statistical models that allow for examination of ways in which different items discriminate across respondents with the same value on a trait	Provides a quantitative measure of item functioning; suitable for scale development	Requires data collection; questions considered in isolation	Questionnaire development
	Latent Class Analysis (LCA) (see and )	Statistical models that allow for examination of error rates associated with different items	Provides a quantitative measure of error rates; suitable for comparing different candidate items measuring the same underlying construct	Requires data collection; questions considered in isolation	Questionnaire development

[↑ Back](#)

## References

[↑ Back](#)