IX. Pretesting

Introduction

Pretesting involves a series of activities designed to evaluate a survey instrument's capacity to collect the desired data, the capabilities of the selected mode of data collection, and/or the overall adequacy of the field procedures. Throughout this text we refer to a "pretest" as the collection of all those techniques and activities that allow researchers to evaluate survey questions and/or survey procedures before data collection begins. In contrast, we use the term "pilot study" to refer to pretesting procedures that employ all the procedures and materials involved in data collection -- i.e., a dress rehearsal before the actual data collection begins. Table 1 is a summary of the most commonly used pretesting techniques, such as pilot study, cognitive interviews, focus groups, and behavior coding.

Differences across cultures/countries may or may not exist at any stage of the data collection process. When multiple languages are used in the same survey, pretesting the different language versions is an essential part of ensuring measurement equivalence and testing translations with the target population (see Translation). In addition, it is often not possible to employ the same mode of data collection across countries participating in a cross-national project. This may be due to several factors; cost and infrastructure constraints and literacy issues are the most common. It is important to test in advance the suitability of the selected mode for the survey topic and population (see Data Collection). Pretesting techniques may have limited application in a given context and culture. Research into how strategies may need to be tailored to suit different populations is only beginning to be undertaken systematically.

Guidelines

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

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

   Rationale

   Pretesting is largely cost and time driven. To make the best use of the various pretesting techniques, it should be determined in advance what questions have to be addressed — whether the 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 should be done in each country participating in the research; however, pretesting techniques may vary across countries, depending on testing traditions and availability of resources and trained staff. Even if some or all of the questions have been used in other studies, pretesting for the local context is necessary in order to assess their performance in the mode and question order in the current study, and with the target population.

Procedural steps

- Using Table 1 as an aid, decide what type of pretesting technique(s) will best fit the study's purpose and the culture within which the study will be conducted.

- Assess whether to conduct the pretest(s) in-house or to contract the testing to an outside organization.

- Establish a time schedule that adequately matches the pretesting design, allowing sufficient time to implement any revisions which may be deemed necessary prior to implementing the full study.

- Budget accordingly.

- Plan on expenses related to interviewer/staff training, respondent recruitment, and incentives.

Table 1 presents the most commonly used pretesting techniques. It provides a brief description, strength and weaknesses, and context in which a particular technique is typically used.

Lessons learned

- When conducting cross-national cognitive interviewing projects, care must be taken to ensure that similar procedures are utilized across all countries involved. If the methodology employed is radically different, it becomes impossible to determine whether observed differences across countries are due to differences in the response process, translation, or the conceptual spectrum. It is not safe to **assume** that procedures for conducting cognitive interviews will be the same across all countries. Differences may exist, for example, in the experience of the interviewers, the location of the interviewing, methods used to recruit participants, and approaches to creating the interviewing protocol. Recent work in seven countries (eight languages) has focused on creating a common approach to cognitive interviewing for questions designed to measure health status [15]. To ensure equivalence, all parties involved in the project agreed upon the method to be used for recruiting participants, administering the protocol, and documenting results.

- Pan [16] investigated the efficacy of cognitive testing as a strategy with Chinese respondents. Her investigation identifies challenges and limitations of taking methods developed in one language and culture and directly applying them to another, and points to the need to include consideration of sociolinguistic conventions appropriate to different cultural groups when conducting cognitive interviews.

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 supplement one another and can logically be combined to maximize the efficiency of the pretest design (see Table 1). It is important to take cultural norms/traditions, as well as interviewer characteristics (see Data Collection), 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

- 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 to techniques that pull all aspects of the project into a more comprehensive study.
  - For example, consider a focus group or in-depth interviews for initial questionnaire construction, cognitive interviews for questionnaire development and refinement, and a field pilot study for an overall test of the survey instrument and field procedures. Often, a pilot study with robust samples can be the best way to test the survey instrument as data analyses with sufficient power can be the most effective way to ascertain if the questionnaire is working as intended.

- 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 (also see Translation).

Lessons learned

- In preparation for the shift from a paper-and-pencil instrument to a computer-assisted instrument incorporating a large audio computer-assisted self-interview (ACASI) component, the U.S. Substance Abuse and Mental Health Services Administration (SAMHSA) implemented a comprehensive pretesting plan [8]. The overarching goal of the pretesting was to develop a computerized instrument that could be used easily by both interviewers and respondents to collect accurate and complete data on the sensitive topic of drug usage. It was also essential that any differences in reporting due to the mode change 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 for the ACASI components, to determine the voice to be used for the audio component, and to assess respondents' ability to enter different types of data into the computer (e.g., open-ended responses). Based on results from these laboratory studies, a field test 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 test, a larger field test, incorporating an experimental design, was conducted . Finally, the revised instrument and procedures were fielded in a split-sample comparison with the original paper-and-pencil instrument to allow researchers to assess the impact on the trend lines.

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

Rationale

The selected pretesting procedures may require skills other than those that available interviewers possess. For example, cognitive interviewing requires an interviewing style which is different from traditional standardized interviewing. It is possible for traditional interviewers to conduct cognitive interviews, but they require additional training to ensure they do not approach the cognitive interview as they would a standardized interview. Sufficient time should be allowed to train staff members and develop protocols that correspond to the selected pretest design.

Procedural steps

- Establish standard procedures across countries in:
  - How to select staff members for the pretest.
  - How to train staff members for the pretest.

    - How to best convey the objective of the task (technique dependent).
    - How/whether to standardize the pretesting protocol.
    - What methods will be used to monitor quality (for example, audio recordings, video recordings, interviewer observations, etc.).
    - How results of the pretest will be analyzed (e.g., whether the analysis will be qualitative or quantitative).
    - How to incorporate changes related to the survey instrument and test whether they lead to improvement.
- If 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.
- Use staff members who have some experience on similar tasks to minimize the amount of training and ensure educated feedback about respondent problems and weaknesses in the questionnaire.

Lessons learned

- Training interviewers and structured protocols are important components of pretesting that can help reduce measurement error, as are recordings of the process. In one project, after a series of cognitive interviews with Vietnamese respondents, a Vietnamese team member insisted that the respondents were not able to understand true/false items because these items were not posed as questions. The translation of the items was shown to a Vietnamese co-worker who, however, did not find any problems. When the interview tapes were reviewed, it was discovered that the interviewer behaved inappropriately and read every item without stopping to get the respondent's answers. In some interviews, interviewers even read the skip instructions to the respondents.

4. Conduct the pilot study in the same mode of data collection (interviewer administered or self-administered) as the survey.

Rationale

Interviewer-administered questionnaires involve listening; self-administered questionnaires involve reading. Interviewer-administered questionnaires involve social interaction between the interviewer and the respondent; self-administered do not. Interviewer-administered questionnaires do not require the respondent to navigate and worry about skip patterns; self-administered do.

Interviewer-administered and self-administered questionnaires produce different context effects (e.g., recency and primacy) and provoke different need for socially desirable responding (see Data Collection). In order to determine how well proposed procedures will work in the field, pilot studies should be conducted in the same mode as the final survey.

Procedural steps

- If different modes of data collection are going to be employed across countries, consider pretesting in the respective modes.

- 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., [17]).

- Use the latest version of the questionnaire and, if applicable, respondent materials.

- Use field administration procedures planned for production data collection.

Lessons learned

- 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, even when the survey questionnaire is not translated to a different language. In fact, a change in mode may necessitate changes in wording or changes in design. For example, cognitive testing for the 2001 US Census showed that more redundancy was needed in the instructions to the "respondent race" question for the respondents to be able to follow the "select one-or-more" option in telephone administration [6]. A slightly reworded version of the instructions and question stem 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 [14].

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

Rationale

In order to most effectively pretest the survey instrument or field procedures, conduct the pretest with respondents from the intended target population or, as relevant, a sub-group within the target population. Similarly, 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 real data collection in terms of cooperation, respondent performance, total interview length, questionnaire performance, survey costs, etc.

Procedural steps

For all pretesting techniques:

- Tailor subject/ respondent recruitment to the population of interest.

- Prepare an informed consent form that reflects the goals and risks of the pretest study as needed or when required by the funding organization.

- Select a sample size that is suitable for the chosen pretesting method.

- Monitor pretest participant recruitment to ensure best use of the chosen pretesting method.

For pilot studies:

- Select a sample large enough to provide sufficient statistical power to detect the differences that will answer the research questions identified in your pilot study analysis plan (with suitable allowances for nonresponse, eligibility, etc.).

- Follow the sample selection protocol planned for the final study.

- Monitor the sample selection

Lessons learned

- 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 Hunt et al. [12] demonstrated that the general population may not be a good judge of the quality of survey questions, even when this is the target population. The 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 and “loaded” words. One possible explanation is that all of the respondents had roughly the same low level of survey experience.

6. Pretest the survey instrument or part of it in each country and in each language.

Rationale

Countries/cultures differ by language and/or cultural systems. When possible, the natural flow of the survey instrument should be tested for each culture and language in order to avoid awkward conversational situations, question order with unpredictable culture-dependent context effects, question repetition not intended in the source, or other culture-specific problems.

Procedural steps

- Select staff members who are fluent in the language of the pretest and sensitive to nuances.

- Standardize key components of training procedures across organizations.

- Develop a procedure for how problems will be reported and addressed in a consistent manner between testing sites.

Lessons learned

- Work conducted by the U.S. Census Bureau to develop a bilingual (English/Spanish) decennial census form has 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 [9]. Testing did not directly assess the English questions, as the wording of the English items had already been nearly finalized. As part of one 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 where concepts were unclear for the Hispanic respondents as opposed to questions that were equally unclear for both English and Spanish speakers.

7. **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. Changes to 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 still preserve the measurement equivalence of the survey instrument across countries can be tolerated (see Translation). Any introduced changes should also be pretested to avoid unforeseen errors (also see Survey Instrument Design).

Procedural steps

- Examine the findings of each pretesting technique used and identify the causes of the any problems discovered.
  - Decide in advance what constitutes a problem. For example, the 10%-20% rule is often used in behavior coding to flag questions: if a question is misread or misunderstood by 10%-20% of respondents, then it is considered problematic. It is important to note that the appropriate cutoff for any particular study is often determined from the distribution, based on the way in which codes are assigned.
  - 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.
  - Examine in what situations and with what types of respondents problems occur.

- If a pilot study has been conducted:
  - Review response distributions and item nonresponse for key study variables.
  - Review interview length.
  - For attitudinal and value variables, check whether items group together as intended in the survey (e.g., perform confirmatory factor analysis).
  - Solicit and review feedback from interviewers.

- 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 results and propose meaningful changes.

- If changes are introduced to the questionnaire or design procedures, plan for another pretest.

Lessons learned

- Pretesting techniques and the results they yield are meaningful only when the selected procedures are culturally appropriate. Not all pretesting techniques have been equally tested and studied across countries; thus, some may not be successfully implemented and lead to meaningless results in certain cultures. For example, studies in psycholinguistics have demonstrated difference in cognitive tendencies among Chinese and English speakers to use counterfactual reasoning [3]. When asked what their thoughts would have been on a 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.

8. 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 the same study.

Procedural steps

In a manner consistent across countries:

- Document the pretest sample selection and recruitment method, including the sample frame and sample size.

- Document the geographical location of the pretest.

- Document respondent characteristics.

- Documents mode(s) of pretest administration.
- Document dates of data collection and organization(s) conducting the interviews.
- Document types of staff conducting pretest (e.g., experienced interviewers, supervisors).
- Document all materials used in the pretest.
- Describe pretest findings and their implications.
- Document any changes made to the survey instrument and the pretesting source that lead to these changes.
- Document the number and types of pretest.

Lessons learned

- 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 their 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), but also by question type (e.g., objective characteristics, behavioral reports, attitudes), response category type (e.g., yes/no, open-ended, quantity), and response error type (e.g., problems with terms, recall problems). 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 topic areas 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 research reports to the catalogue to strengthen the utility of the site.

Table 1. Pretesting methods, their strengths, and weaknesses.

	Pretesting Method	What it is	Strengths	Weaknesses	Most Common Use
Field Methods	Field pilot study (for an overview, see [10])	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	very 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 [9])	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);	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., [13]; also, [10])	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 consisted use of the coding scheme; does not identify the exact problem in a question with many codes	questionnaire testing; field management

	Focus groups (see [6] for an overview; also [10])	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 in the intended direction	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 follow up; requires small sample 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
Cognitive Laboratory Methods (for an overview, see [9])	Vignettes (e.g., [19])	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., [3])	questionnaire development; concept understanding test
	Concurrent think-aloud [2] [6].	Respondents' report 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 provide unrealistic picture of how respondents answer questions in the field; coding may be burdensome; assumes respondents are able to indentify and report what information they used to come up with a response to the survey question; respondents may begin to over interpret the questions and come up with problems that do not exist in the natural context	questionnaire development

	Retrospective think-aloud [1]	Interview with respondents after they have completed a survey about how they came up with answers to specific questions	does not interfere with the response formation process	assumes respondents are able to indentify and report what information they used to come up with a response to the survey question; assumes information is still available in short-term memory	questionnaire development
Other	Expert review (for an overview, see [10])	Review of draft materials by experienced methodologists, analysts, 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	questionnaire development
	Question Appraisal System (for example, [23])	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	questionnaire development
	Usability Testing [11] [22]	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 cheap - can be conducted with employs of the survey organization; requires small sample sizes	time consuming;	field work test

Statistical Modeling	Multi-trait-multi-method Database [20]	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 intense; questions are considered in isolations, so question order effects might be ignored	
	Item Response Theory Approach [18]	Statistical models that allow to examine how 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	

Glossary

Audio computer-assisted self-interview (ACASI)	A mode in which the respondent uses a computer which displays the question on screen and plays audio recordings of the questions to the respondent, who then enters his/ answers.
Behavior coding	Systematic coding of the interviewer-respondent interaction in order to identify problems that arise during the question-answer process
Cognitive interviews	A pretesting method designed to uncover problems in survey items by having respondents think out loud while answering the question.
(Cognitive) vignettes	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 .
Context effects	The impact of question context, such as the order or layout of questions, on survey responses.
Coordinating center	A research center that facilitates and organizes cross-national research activities.
Double-barreled (questions)	Survey questions that inadvertently ask about two objects at once.
Field pilot study	Small scale rehearsals of the data collection conducted before the main survey.
Focus group	Small group discussions under the guidance of a moderator, often used in qualitative research, that can also be used to test survey questionnaires and survey protocols.
Item nonresponse/ item-missing data/ missing data	The lack of information on individual data items for a sample element where other data items were successfully obtained.
Loaded questions/words	Questions that are worded in such a way that invite respondents to respond in a particular way.

Measurement equivalence	Equivalence of the calibration system used in the questionnaire and the translation.
Mode	Method of data collection.
Pilot study	A pretesting technique that involves all procedures and materials that will be involved in data collection; a dress rehearsal before the actual data collection begins.
Pretesting	A collection of techniques and activities that allow researchers to evaluate survey questions and/or survey procedures before data collection begins.
Primacy	Context effects in which the placement of the item at the beginning of a list of response options increases the likelihood that it will be selected by the respondent.
Recency	Context effects in which the placement of the item at the end of a list of response options increases the likelihood that it will be selected by the respondent.
Response distributions	A description of the values and probabilities that a particular response was selected.
Sampling frame/sample frame	Lists or materials used to identify all elements (e.g., persons, households, establishments) of a survey population from which the sample will be selected . These lists or materials can include maps of areas in which the elements can be found, lists of members of a professional association and registries of addresses or persons.
Target population	The finite population for which the survey sponsor wants to make inferences using the sample statistics.

References

[1] Belson, W. A. (1981). *The design and understanding of survey questions*. Aldershot, UK: Gower.

[2] Bickart, B.& Felcher, M. (1996). Expanding and enhancing the use of verbal protocols in survey research. In N. Schwarz & S. Sudman (Eds.), *Answering Questions: Methodology for Determining the Cognitive and Communicative Processes in Survey Research*. San Francisco: Jossey-Bass, 115-142.

[3] Bloom, L. (1981). *The linguistic shaping of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[4] Bolton, R., & Bronkhorst, T. (1996). Questionnaire Pretesting: Computer-Assisted Coding of Concurrent Protocols. In N. Schwarz & S. Sudman (Eds.), *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*. San Francisco: Jossey-Bass.

[5] Davis, D. K., Blair, J., Crawley, E. L., Craig, K. M., Rappaport, M. S. B., Baker, C. A., & Hanson, S. L. (2001). *Census 2000 quality survey instrument pilot test*. Report submitted by Development Associates, Inc., Arlington, VA. Washington, DC: US Census Bureau, Center for Survey Methods Research.

[6] Davis, W., & DeMaio, T. (1993). *Comparing the think-aloud interviewing technique with standard interviewing in the redesign of a dietary recall questionnaire*. Alexandria, VA: ASA Section on Survey Research Methods.

[7] Fowler, F.J. Jr. (1995). *Improving survey questions*. Vol. 38, *Applied social research methods series*. Thousand Oaks, CA: Sage Publications.

[8] Gfroerer, J., Eyerman, J., & Chromy, J. (Eds.) (2002). *Redesigning an ongoing national household survey: Methodological issues*. DHHS Publication No. SMA 03-3768. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

[9] Goerman, P., Caspar, R., Sha, M., McAvinchey, G., & Quiroz, R. (2007). *Census bilingual questionnaire research, Final round 2 report.* Retrieved June 3, 2008, from http://www.census.gov/srd/papers/pdf/ssm2007-27.pdf

[10] Groves, R., Fowler, F. Couper, M. Lepkowski, J., Singer, E. & Tourangeau, R. (2004). *Survey methodology*. Hoboken, NJ: Wiley & Sons.

[11] Hansen, S.E., and Couper, M. (2004). Usability Testing as a Means of Evaluating Computer Assisted Survey Instruments. In S. Presser et al. (Eds.), *Methods for Testing and Evaluating Survey Questionnaires*. New York: John Wiley & Sons.

[12] Hunt, S.D., Sparkman, R. D., & Wilcox, J. B. (1982). The pretest in survey research: Issues and preliminary findings. *Journal of Marketing Research 19*, 269-273.

[13] Mangione, T., Fowler, F. J., & Oksenberg, L. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics 8(3),* 293-307.

[14] Martin. E., & Gerber, E. (2004). Results from recent methodological research on the Hispanic origin and race questions. Paper presented to the Population Association of America.

[15] Miller, K., Fitzgerald, R., Caspar, D., M., Gray, M., Nunes, C., Padilla, J.L., et al. (2008). Design and analysis of cognitive interviews for cross-national testing. Paper to be presented at the International Conference on Survey Methods in Multinational, Multiregional and Multicultural Contexts, Berlin, Germany.[1]

[16] Pan, Y. (2004). Cognitive interviews in languages other than English: Methodological and research issues. *Proceedings of the American Statistical Association*, *Section on Survey Research Methods*.

[17] Redline, C., R. Smiley, M. L., DeMaio, T., & Dillman, D. (1999). Beyond concurrent interviews: An evaluation of cognitive interviewing techniques for self-administered questionnaires. Proceedings of the American Statistical Association, Section on Survey Research Methods.

[18] Reeve, B., & Mâsse, L. (2004). Item Response Theory (IRT) Modeling for Questionnaire Evaluation. In S. Presser et al (Eds.), *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley.

[19] Rossi, P., & Anderson, A. (1982). The factorial survey approach: An introduction. In P. Rossi & S. Nock (Eds.), *Measuring social judgments: The factorial survey approach*. Thousand Oaks, CA: Sage.

[1] The U.K. fieldwork was supported by a grant from the Economics and Social Research Council

[20] Saris, W., van der Veld, W., & Gallhofer. I. (2004). Development and Improvement of Questionnaires Using Predictions of Reliability and Validity. In S Presser et al (Eds.), *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley.

[21] Snijkers, G. (2002). *Cognitive laboratory experiences on pretesting computerized questionnaires and data quality*. Heerlen: Statistics Netherlands.

[22] Tarnai, J., & Moore, D. (2004). Methods for Testing and Evaluating Computer-Assisted Questionnaires. In S Presser et al (Eds.), *Methods for Testing and Evaluating Survey Questionnaires*, New York: Wiley.

[23] Willis, G. B., & Lessler, J. T. (1999). Question appraisal system BRFSS-QAS: A guide for systematically evaluating survey question wording. Report prepared for CDC/NCCDPHP/Division of Adult and Community Health Behavioral Surveillance Branch.