

Standards for Pretesting Questionnaires and Survey Related Materials for U.S. Census Bureau Surveys and Censuses

Theresa DeMaio
Census Bureau, USA

1. Introduction

The U.S. Census Bureau has, over the last three years, developed a series of standards, principles, and guidelines that cover a wide variety of topics to standardize and ensure the quality of the Census Bureau's products. One that is particularly suited to the goals of QUEST is the standard for pretesting questionnaires and related materials for surveys and censuses. At one of the very first QUEST meetings, developing standards for pretesting across the member statistical agencies was viewed as one of the goals of the group. My paper describes the Census Bureau's attempts to develop such standards.

The Census Bureau had a policy in place since 1995 that applied to its demographic area, which conducts household surveys. The standard I will be talking about today expands that policy to pertain to all the censuses and surveys the Census Bureau conducts.

Developing the standard was a time consuming process. An interdivisional committee consisting of the authors, represented the decennial, demographic, and economic, and research areas of the Census Bureau. We held numerous meetings to discuss the issues involved in pretesting, what should be included as part of a standard, and how differences between the surveys conducted by different areas of the Census Bureau should affect what is included in the standard. At the same time as we considered what the optimal components of the standard should be, we also had to keep in mind the practical limitations imposed by being able to put the standards into practice. After many drafts and discussions with managers in the survey operations areas, a standard was approved by the executive staff in July, 2003.

The standard includes a minimal standard and three recommended standards. The minimum standard is required, but the three recommended standards apply to special types of data collections, and are not currently required. Their status may change, however, as the standard is reviewed over time. Current plans are to review the standards after 5 years. The standard also includes an attachment that describes a variety of pretest methods. In this paper, we present the content of the standard, outline the different pretesting methods, and give an example of how the standard is applied. Interested readers can download a copy of the standards from the following web address: <http://www.census.gov/srd/pretest-standards.pdf>.

2. Minimum Standard

So... what is the minimum standard? The minimum standard requires testing that exposes respondents to the questionnaire and shows that the questionnaire "works." Whether a questionnaire works is kind of ambiguous; there are no concrete statistical criteria that must be met. The theory is that by getting respondents to try out draft versions of the question, we learn about potential data problems and have an opportunity to correct them before they are encountered in the field.

Evidence that a question works can include the results of pretesting conducted by the Census Bureau, the survey sponsor, or a contractor. It can also include the results of research demonstrating that the use of the question in a prior survey worked. Not just that it was used, but that it worked.

This minimum standard applies to testing English-language questionnaires as well as all foreign languages that will be used in data collection. The Census Bureau currently has guidelines for translating questionnaires into foreign languages and we are in the process of developing guidelines for pretesting questionnaires in foreign languages. In the meantime, the procedures that are used for English-language questionnaires are being adopted for testing of questionnaires in foreign languages.

The minimum standard also applies to all questionnaire versions when multiple modes of questionnaire administration are conducted. Often meaningful changes, such as changes to the wording or format of the questions, are made to reflect mode-specific functional constraints or advantages. According to the standard, each version must be tested to facilitate maximum consistency in the interpretation of question purpose across modes, despite the structural or presentation differences.

3. Recommended Standards

As noted previously, there are three recommended standards, which apply to surveys and censuses with special circumstances. The first is a recommended standard for data with major policy implications, which we define as key economic or socio-economic indicators. The standard contains a list of what these indicators are; for example, the unemployment rate and the poverty rate. It recommends that split panel testing be undertaken whenever changes are made to the questionnaires, since changes in survey questions or procedures may affect the continuity of time series data. This allows the Census Bureau to isolate the effects of real changes in the statistics over time from changes due to alterations in the questions, survey design, or mode. This may not be cost-effective for other surveys, because the large costs involved in mounting a split panel test can only be justified for surveys with important policy implications.

The second recommended standard is for supplemental instruments and materials. The standard recommends testing of such things as advance letters, reminder postcards, supplemental instruction sheets, and letters mailed with replacement questionnaires. The focus here is on the decennial census of population and housing, because of its high profile and the large impact that procedural failures have. It is important to test all the pieces of correspondence to respondents, or to base letters and instructions on prototypes or boiler plate language that has been pretested.

It is also important to view all the components of a mailout to respondents as a whole, to prevent errors visible to the general public and the media that might reflect badly on the Census Bureau. This includes testing the outgoing envelope and all its components, depending on what is relevant. It could be communication in advance of the questionnaire mailout or the self-administered questionnaire package itself. In addition to communications in the census itself, the second recommended standard also advocates testing of data collection instruments for pre-census operations such as permanent address listing, block canvassing, and group quarters frame development.

The third recommended standard is for electronic self-administered questionnaires. It recommends that the entire programming operation be tested, not just the form. This includes such things as the log-in or access mechanisms used by respondents, interfaces for navigating between screens, the performance of the automated edit functions, the design of the help screens, and the utility of the submission procedures. Such broad testing will uncover usability problems related to the platforms used by respondents, the design of the screens, navigation through the instrument, and bugs in the software. Some program areas of the Census Bureau have tested and adopted a style guide for creating electronic instruments. As long as the style guide has been tested, interfaces designed according to its criteria do not need to be pretested.

4. Synopsis of Pretesting Methods

In addition to stating the rules to be followed in pretesting, the standard attempts to familiarize employees with the methods that can be used for pretesting. The standard includes a 9-page description of a variety of pretesting methods, including both qualitative and quantitative methods, and methods used during questionnaire development (which we call pre-field methods) as well as field testing. This document is meant to be broad in scope, including methods that are commonly used only for economic surveys, and describing the demographic and economic uses of others.

The general rule for whether a pretesting method meets the standard is whether it involves some kind of testing with respondents. This excludes expert reviews, in which questionnaires are reviewed by subject matter experts or survey methodologists for potential problems. This was a deliberate decision; the questionnaire needs to be tested with respondents to be able to show that it works. The standard explicitly states that expert reviews are allowed only under extreme time pressure, although it is described in the document.

Figure 1 includes a list of the methods described in the document, broken down separately for pre-field techniques (those that are used during the development of the survey questionnaire) and field techniques. It shows whether each one is qualitative or quantitative, and whether it is used for demographic surveys, economic surveys, or both. A brief description of each of the methods listed follows.

Respondent focus groups are used in both demographic and economic surveys to gather information about a topic before questionnaire construction starts. Groups of between 8 and 12 respondents are led by a trained facilitator to elicit information about how respondents structure their thoughts about a topic, how they understand general concepts or specific terminology proposed for the survey, and whether they think proposed survey content is sensitive.

Exploratory studies tend to be used differently in demographic and economic surveys. In both cases they involve unstructured interviews between researchers and potential respondents in preliminary stages of questionnaire development. But in demographic surveys, they focus on potential respondents' understanding of terms, concepts, and cultural influences, while in economic surveys, they focus on the match between desired data and available data in company records.

Cognitive interviews consist of one-on-one interviews using a draft questionnaire in which respondents are asked to provide information, either during the interview or after the questionnaire is completed, about how they interpreted the questions and how they arrived at their

answers. Demographic and economic questionnaires use the same general type of cognitive interview process.

Usability techniques are one-on-one encounters between researchers and respondents that focus on the design elements of electronic questionnaires, such as language, fonts, icons, and layout. They often reveal information about the question wording as respondents navigate through the questionnaire.

Methodological expert reviews are conducted by survey methodologists or questionnaire design experts rather than subject matter experts, and their objective is to evaluate the questionnaire for potential respondent and interviewer task difficulty. The asterisk in Fig. 1 denotes that, as I noted before, this method used alone is not sufficient to meet the pretesting criteria.

Behavior coding applies only to demographic surveys because it is an interview-focused technique, and the Census Bureau conducts its economic surveys by mail. It involves systematic coding of the interaction between interviewers and respondents from live or taped telephone or field interviews to collect quantitative information. The behavioral aspects that are coded are ones that are associated with questionnaire problems, such as interviewers misreading the question or respondents requesting clarification after the interviewer asks the question.

Respondent debriefing involves using a structured questionnaire following data collection to elicit information about respondents' interpretations of survey questions. These can be quantitative if precoded categories are used in the questionnaire or qualitative if extensive open-ended questions are asked. For demographic surveys these additional questions are generally added at the end of the telephone or personal visit interview; for economic surveys, they are more likely to take the form of a supplemental questionnaire included in the mailing package.

Interviewer debriefing is another method that only applies to interviewer-administered questionnaires, and thus only to demographic questionnaires at the Census Bureau. It involves using the accumulated knowledge of survey interviewers as they administer the questionnaires to provide insight into questionnaire problems. As the asterisk in Fig. 1 indicates, its use alone is not sufficient to meet the pretest criteria since it does not include direct contact with respondents.

Analysts' feedback is a method unique to the economic area. It involves review of records compiled by program staff analysts as they conduct their routine survey responsibilities of following up with respondents to investigate suspicious data flagged by edit failures, or responding to inquiries by respondents who are phoning in for help. Because the method itself does not involve contact with respondents, it does not meet the pretest criteria.

Split panel tests refer to controlled experimental testing of questionnaire variants or data collection modes to determine which one is better or to measure differences between them. This is appropriate for both demographic and economic surveys, and allows the effects of changes in the instruments being varied to be separated out from the effects of real change over time.

Analysis of field pretest data refers to analysis of item nonresponse rates, imputation rates, edit failures, and response distributions from the data collected during the field test. This is obviously all quantitative data, and the types of measures calculated may differ between demographic and economic surveys. For example, edit failure rates are more frequently reviewed in economic surveys than demographic surveys, while the opposite is true for item nonresponse rates.

5. Use of the Minimal Standard

Most of these methods can, under the right circumstances, meet the pretesting standard. However, in reality, pre-field methods, and most frequently cognitive interviews, are most likely to meet the strict deadlines that generally exist for making changes in production surveys. That being the case, we next present an application of the use of the minimal standard.

The Center for Survey Methods Research in the Statistical Research Division recently conducted cognitive testing of newly proposed identity theft questions for the National Crime Victimization Survey. Initial testing of this series of questions showed several problems.

Respondents were asked a series of yes-or-no questions about whether someone used or tried to use a household member's credit card or credit card number without permission, whether someone used or tried to use any existing accounts other than a credit card, and whether someone used or tried to use personal information to run up new debts. After these questions, respondents were instructed to answer the remaining questions about the most recent one of these episodes. We learned from respondents that this was simply not always possible, since the incidents themselves may have involved multiple types of issues. Another question asked whether the respondent knew the person who misused the information. Although the intent was to see if the respondent ever found out who did it, that was not always the response provided. Most respondents thought we were asking if they knew the person personally, and so they misreported. These problems were so blatant that the questions were revised after a few interviews were conducted. Subsequent testing was conducted, and after the third round of interviews, the problems seemed to have been eliminated and no new problems identified.

At the end of the cognitive testing, staff prepared a report containing the results, recommendations for questionnaire revision, and discussions of other issues that arose during the testing. A joint meeting between the sponsor, the Census Bureau's Crime Statistics staff, and CSMR staff who conducted the pretesting was held to discuss the issues, the recommendations, and finalize the questionnaire. In the vast majority of cases, the recommendations were either accepted as is or in a modified form. In this way the standard can be used to provide evidence that the questions work before they are fielded.

Pre-Field Techniques

Respondent Focus Groups	Qualitative	Demographic/ Economic
Exploratory or Feasibility Studies	Qualitative	Demographic/ Economic
Cognitive Interviews	Qualitative	Demographic/ Economic
Usability Techniques	Qualitative	Demographic/ Economic
Methodological Expert Review *	Qualitative	Demographic/ Economic

Field Techniques

Behavior Coding of Respondent/ Interviewer Interaction	Quantitative	Demographic
Respondent Debriefing	Quantitative/ Qualitative	Demographic/ Economic
Interviewer Debriefing *	Qualitative	Demographic
Analyst's Feedback*	Qualitative	Economic
Split Panel Tests	Quantitative	Demographic/ Economic
Analysis of Item Nonresponse Rates, Imputation Rates, Edit Failures, and Response Distributions	Quantitative	Demographic/ Economic

* Does not meet the eligibility criteria for pretesting standard

Figure 1. List of Pretesting Methods Included in Pretest Standard