

Experiences Implementing Establishment Survey Questionnaire Development and Testing at Selected U.S. Government Agencies

Karen L. Goldenberg, U.S. Bureau of Labor Statistics¹

Amy E. Anderson and Diane K. Willimack, U.S. Census Bureau²

Stanley R. Freedman and Robert H. Rutchik, U.S. Energy Information Administration¹

Luann M. Moy, U.S. General Accounting Office¹

Amy E. Anderson, U.S. Census Bureau, Room 3110-4, Washington, DC 20233

amy.e.anderson@census.gov

I. Introduction

A defining feature of the United States Federal statistics system is the decentralization of statistical survey activities among multiple agencies. Consequently, there is no overall approach to questionnaire development, evaluation, and testing (QDET) for establishment surveys among government agencies. Besides being subject to legal restrictions on data sharing, these agencies have different data collection goals and, to some degree, different target sub-populations within the overall population of establishments. As a result, methods for developing and testing establishment survey questionnaires have evolved somewhat independently.

The purpose of this paper is to identify and compare methods for developing and testing establishment survey questionnaires across four U.S. Federal agencies: the Bureau of Labor Statistics, the Census Bureau, the Energy Information Administration, and the General Accounting Office. In the paper, we raise common QDET issues for the establishment setting and explain notable differences among agencies.

A. The Authors

The authors of this paper are survey methodologists in our respective agencies. We are not subject matter experts, nor do we manage the surveys for which we provide expertise. Rather, we function as methodological consultants for our agencies' subject matter experts and survey operations staffs. This paper reports on our own experiences in this role. Further, the paper does not represent a comprehensive survey of all the QDET activities that go on in our organizations, but instead reflects our knowledge of agency activities. In addition, the activities described do not represent official agency policies, but in many cases are responses to data needs that have arisen over time.

¹ The opinions expressed in this paper are those of the authors and do not reflect official policy of the Bureau of Labor Statistics, Energy Information Administration, or the U.S. General Accounting Office.

² This paper reports the results of research and analysis undertaken by Census Bureau staff and their collaborators. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

B. Organization of the Paper

This paper is organized into four sections. Section I, the introduction, describes the purpose of the paper, outlines its organization, and provides a brief picture of the participating organizations. Section II lays out the process by which these agencies perform questionnaire development and testing. Section III presents a description of the techniques the four agencies use in questionnaire development, testing, and evaluation, including examples of how the agencies implement the techniques. Section IV summarizes the common threads among the four agencies and discusses the reasons for the differences.

C. Participating Agencies

The **Bureau of Labor Statistics (BLS)** is an agency in the Department of Labor in the executive branch of the U.S. government. It is the principal fact-finding agency for the Federal Government in the broad field of labor economics and statistics. BLS establishment surveys collect information about business establishments, their employees, and their products or services. These surveys are the source of data on payroll, employment, occupational employment by industry, pricing indexes, employer costs, and occupational safety. Most of these surveys provide data for ongoing data series, with sample establishments reporting data monthly or quarterly, in some cases for many years, or indefinitely. Others are annual surveys. Single time surveys are unusual, but may be conducted to get a better understanding of a specific situation or data issue. Several of the continuing BLS establishment surveys are collected by designated State agencies under a Federal-State cooperative program, using forms and procedures developed by BLS.

The **Census Bureau** is an agency in the Department of Commerce in the executive branch of the U.S. government. It is a key collector and provider of data about the United States population and economy. In addition to conducting the decennial census of the population and numerous other demographic surveys, the Census Bureau collects data describing businesses, institutions and governments. The Census Bureau conducts more than 100 different annual, monthly, or quarterly economic surveys, along with an economic census in years that end in 2 and 7. Monthly indicator surveys collect data such as: housing starts; merchandise trade; manufacturing shipments, orders, and inventories; retail sales; wholesale sales; business sales and inventories; and corporate profits. Annual surveys collect data such as sales, assets, payroll, capital expenditures, and employment. The economic census collects official measures of output for all industries at detailed geographic levels.

The **Energy Information Administration (EIA)** is a statistical and analytical agency within the U.S. Department of Energy, also in the executive branch. EIA provides independent data, forecasts, and analyses to promote sound policy-making, efficient markets, and public understanding regarding energy and its interaction with the economy and the environment. EIA fields over 50 different surveys of business establishments every year. These surveys collect information on the supply, stocks, transportation, distribution, sales, and consumption of petroleum, natural gas, electricity, coal, nuclear, and alternative fuels.

The **U.S. General Accounting Office (GAO)**, unlike the other three agencies, is an agency of the U.S. Congress, in the legislative branch of the U.S. government. The GAO is also not a statistical agency. Congress asks GAO to study the programs and expenditures of the federal government and how the federal government spends taxpayer dollars. GAO advises Congress and the heads of executive agencies (such as the Environmental Protection Agency, Department of Defense, and Health and Human Services) about ways to make government more

effective and responsive. GAO evaluates federal programs, audits federal expenditures, and issues legal opinions, conducting sample surveys to support statistical analyses for these efforts, and reports its findings and recommendations to Congress.

II. Designing and Redesigning Establishment Survey Data Collection Instruments: A Process Perspective

The development, testing, and evaluation of data collection instruments for establishment surveys represent a process. The different phases of this process occur along a continuum, consisting of separate sets of linked activities. These are:

- *Questionnaire development and testing (QD&T)*. This phase of the process includes the activities leading up to the design of the questionnaire. In this context, "testing" refers primarily to pretesting – that is, qualitative or quantitative research with small and non-generalizable numbers of respondents. Results are used to make changes to the data collection process or instrument.
- *Questionnaire evaluation (QE)*. As used here, questionnaire evaluation occurs *after* an instrument has been used to collect data. In our view, questionnaire evaluation takes place following a pilot test or the first wave of collection for a new survey, or at any point where data from an ongoing survey have been reviewed. The resulting data from an evaluation are fed back into QD&T if it appears that changes are warranted.

The phases along the continuum are not always distinct, often blending into one another. In addition, the process is not linear but iterative, characterized by numerous feedback loops, where information obtained at any point in the process can be used to modify and update draft questionnaires or existing plans ("review and revise"). The sequence of activities within the process is not fixed. Different agencies use many of the same approaches to QDET, but not necessarily in the same order or at the same place in the process.

A. Steps in the Questionnaire Development and Testing Process

The QD&T process begins with identifying requirements for the survey—that is, the objectives, the concepts, the measures to be obtained, and other aspects of survey content. This activity relies heavily on needs of survey sponsors and on interactions with various groups of stakeholders. Once the requirements have been determined, staff responsible for questionnaire development turn the concepts into specific measures and draft the questionnaires or data collection instruments. Questionnaires are then discussed with experts in instrument design and with subject matter experts, pretested with respondents, and revised based on the outcomes of the preceding steps. These steps are then repeated as needed or as permitted by the data collection schedule.

1. Develop Requirements for the Questionnaire

Determining survey requirements is the first step in the process of questionnaire development in each of our agencies. There are two types of requirements. The first group comprises the content for the survey—that is, the data the agency will collect in the new or redesigned questionnaire. The second group consists of basic design constraints and non-content aspects of the instrument, such as limitations on length, the positioning of mailing information, or formatting for automated data capture.

Determining survey content is often done in consultation with or by obtaining information from staff within one's agency, from other organizations inside and outside the government, and from potential respondents. Examples of this type of information include understanding how an industry works, whether establishments maintain the specific data requested, whether the data correspond to the underlying concept, and whether the terminology is familiar to respondents. These consultations often lead to changes in questions, especially if their results demonstrate a mismatch between planned concepts and available data.

Survey sponsors and other stakeholders often participate in the development of survey content requirements, and/or play an active role in reviewing the results of meetings with respondents, internal and external subject matter specialists, and others. Each of these reviews functions as a feedback loop to the development path. This process has two important consequences. First, if changes are needed to requirements, the various parties that share in the review will understand the need for them. As a result, they will be more likely to agree to those changes. The second consequence is that the interested parties have some investment in the survey and agree to support or encourage participation.

2. Draft data collection instruments/approaches

The second step in the QD&T process is to design the draft data collection instruments. In many cases this is a team activity. One person may write the questions or devise the format, but sponsors, stakeholders, and other relevant groups (e.g., procedures or systems staffs) participate actively as well. It is at this phase of the process where requirements other than content come into play. Mailing requirements, processing requirements, automated data capture, and similar specifications affect the appearance of the instrument. Likewise, agency style guidelines affect the design of electronic forms. These requirements influence question sequence, question wording, and the layout of self-administered questionnaires.

3. Subject the instrument to expert review

In general, expert review is a process by which experienced and knowledgeable content or methods experts apply a systematic approach to evaluating a topic or content area. It plays an important role in the QD&T process. Individuals with expertise in questionnaire design review the instrument for question content, question flow, and the ease with which the questions are likely to be understood. In addition, if the questionnaire was designed for self-administration (on paper or on screen), the reviewer will look at consistency of formatting and placement of questions and response frameworks. Similarly, an expert review of an interview script or a touchtone telephone self-response script looks at the clarity of questions, and for the presence of any unnecessary words, instructions or other question elements that might inhibit response.

Subject matter experts, including the survey sponsors, also participate in an expert review. They generally focus on the content of questions, instructions, definitions, terminology, and other aspects of the topic to be sure that the questionnaire covers the subject completely and

clearly. Systems specialists may also review a draft instrument to verify that it meets the needs of the processing staff's automated operations.

Following these reviews, the questionnaire designer incorporates the feedback into the questionnaire. In some cases, doing so requires negotiating competing needs or contradictory feedback. The expert review cycle may be repeated several times until all parties are satisfied.

4. Pretest the instrument with respondents

An important part of the QD&T process is pretesting with respondents. No matter how well subject matter experts think they know the population being studied, it is impossible to anticipate how respondents will react to and respond to questions. Pretesting allows the developers to see actual or potential limitations and problems in respondent comprehension, question wording, and form layout, as well as to identify discrepancies between requested information and data in establishment records. Results may also suggest solutions to question problems. Following the pretest, the instrument designers review the results and provide recommendations to sponsors and other participants in the QD&T process. Again, negotiation may be required, e.g., if pretest results are internally inconsistent, contradict staff expectations, or adversely impact systems specifications. Pretesting may be repeated as schedules and budget allow. Ideally, it continues until each new interview or test reveals no new information, and all known problems have been resolved.

5. Revise and Review

As noted earlier, QD&T is an iterative process. All of the above steps, in sequence or individually, may be repeated one or more times. The process of drafting and redrafting an instrument may be ongoing, until it is finally submitted for expert review; then it may begin again. Likewise, pretesting that reveals important flaws will probably result in revisions to the questionnaire. Sometimes pretesting points to problems with the concepts or measures planned for the survey, at which point underlying concepts and requirements may need to be revisited.

B. Steps in the Questionnaire Evaluation Process

We consider questionnaire evaluation (QE) to be the part of the QDET process that follows data collection. Regardless of whether the data collection source is a pilot study or an operational survey, the collected data are examined carefully for characteristics related to data quality. While data quality may be affected by a number of factors, it may also be considered an indicator of the performance of the data collection instrument.

Like the QD&T portion of the process, QE involves feedback loops. Typically, feedback from QE informs QD&T. Since pilot studies require commitment of substantial resources, the results are used to revise questionnaires and to improve the quality of the collected data. In contrast, the primary goal of reviewing data collected in ongoing or recurring surveys is to assess data quality. Only to a lesser degree does such a review result in revisions to the questionnaire. When potential problems are identified as a result of QE, they should be considered through the QD&T process before actually changing the questionnaire.

III. Techniques Used in Establishment Survey QDET

Section II described the iterative process by which the agencies perform QD&T and QE for establishment surveys—that is, *what* we do. In Section III we discuss the approaches, techniques, activities, and methods the agencies use during the QDET process—that is, *how* we do it. We identified 16 different techniques that we have used to conduct QDET. Each technique is used by at least one of our agencies.

Table 1 shows the methods and the QD&T and QE processes. More specifically, the left-hand column of Table 1 lists the methods we identified; we describe each of these methods below. The remaining columns of the table show the steps of the process, present the authors' assessments of whether our agencies use each method, and if so, the step(s) in the QDET process where the method is applied.

The material in this section is presented with two caveats: First, the methods and approaches described here and attributed to an agency have been used by that agency but not necessarily for every questionnaire or establishment survey designed or revised by that agency. Second, the implementation of a method as described here reflects the knowledge and experience of the authors. We do not claim to present a complete picture of any agency's QD&T and QE activities for establishment surveys.

Table 1.
Steps and Methods of Questionnaire Development/Testing and Evaluation Process: Agency Activities

METHODS \ STEPS	Questionnaire Development and Testing				Questionnaire Evaluation
	Develop Requirements (1)	Draft Data Collection Instruments (2)	Subject Instrument to Expert Review (3)	Pretest with respondents (4)	Evaluate Collected Data (5)
Consulting interested parties	BLS, CB, EIA, GAO	BLS, CB, EIA, GAO	BLS, CB, EIA, GAO	BLS, CB, EIA, GAO	
Conducting Focus Groups	BLS, EIA, GAO	BLS		GAO	
Conducting Pre-Survey Design Visits	BLS, CB, EIA, GAO				
Integrating Content/Noncontent Questionnaire Elements	BLS, CB, GAO	BLS, CB, EIA, GAO			
“Borrowing” Questions from other Questionnaires	BLS, CB, EIA, GAO	BLS, CB, EIA, GAO	CB, EIA		
Conducting methodological Expert Review			BLS, CB, EIA, GAO		
Cognitive Pretest Interviewing	BLS			BLS, CB, EIA, GAO	
Testing using Vignettes		CB		BLS, CB	
Questionnaire pretesting				BLS	
Functionality and Usability Testing	BLS, CB	BLS, CB, EIA	BLS, CB, EIA, GAO	BLS, CB, EIA, GAO	
Conducting pilot/field test					BLS, CB
Analyzing collected data	BLS	BLS			BLS, CB, EIA, GAO
Split sample testing					BLS, CB
Debriefing survey staff and interviewers	BLS, EIA	BLS			BLS, CB, EIA
Debriefing respondents					BLS, CB
Reinterviewing respondents					CB, GAO

Agencies: BLS – Bureau of Labor Statistics; CB – Census Bureau; EIA – Energy Information Administration; GAO – General Accounting Office

A. Techniques and Approaches used in Questionnaire Development and Testing

1. Consulting Interested Parties: Sponsors, Stakeholders, and Subject Matter Experts

The Method. A key element of QD&T involves consultation with the parties having an interest in the survey. For purposes of this paper we have divided these interested parties into three groups: survey sponsors, stakeholders, and internal or external subject matter experts.

Survey sponsors are the groups or organizations that actually fund a survey. Sponsors may be external to the agency conducting the survey, or internal agency program offices. Survey sponsors wield a great deal of control over survey content and questionnaire design, since they provide the funding and represent a constituency whose data requirements *must* be satisfied.

Stakeholders are groups or individuals who have an interest in how the survey is designed or fielded, such as users of the data products that come from the survey. Stakeholders can include key survey respondents, trade or professional groups that represent data providers or data users, survey methodology groups concerned with data quality, and other interested parties. These groups, usually external to the organizations sponsoring or conducting the survey, can make important and influential contributions to survey content and methods.

Subject matter experts also play a critical role in the design, testing and evaluation process. External subject matter experts are typically data users. They suggest data needs and provide feedback regarding the appropriateness of measurements relative to those needs, often addressing substantive aspects of data quality. Internal subject matter experts come from within an agency, and have a substantive interest in the data. Subject matter experts and stakeholders may overlap.

In many cases, internal subject matter experts act as internal sponsors for the survey. These individuals work in the survey or program offices that conduct surveys funded directly by their own budgets, or they act as agents on behalf of an outside sponsor. The internal subject matter experts in the sponsoring program offices may determine data requirements, propose and construct questions, review questions for conceptual and substantive accuracy, and contribute to data collection and analysis plans.

Survey methodologists in our agencies essentially work as consultants to the internal or external sponsors, providing advice and recommendations on questionnaire construction, testing or survey operations. They work in a collaborative manner, particularly with internal sponsors. Nevertheless, their role is somewhat tenuous. Since sponsors can reject or modify what the consultants suggest, sometimes for reasons that are not related to methodology, methodologists must emphasize evidence supporting their recommendations.

Agency program offices, and in some cases the methodologists themselves, consult with internal and external stakeholders and subject matter experts. A variety of methods may be used for collaboration or consultation. Regardless of the method, however, the participation of sponsors, stakeholders and subject matter experts in the QD&T process is a matter of degree, depending on the position and role of those experts in the governing and policy-making process.

Implementation at our Agencies. Federal funding to conduct surveys constitutes most or all of the agencies' budgets. As a result, while all four agencies consult with survey sponsors, stakeholders and subject matter experts at all stages of the QD&T process, methods for interacting with them differ. At GAO, Congress proposes and requests specific research activities. In the other agencies, this federal sponsorship is manifested in agency program offices and subject areas. In addition, the Census Bureau and, to a lesser extent, EIA receive funding from outside organizations to conduct surveys on their behalf. Most of these external sponsors are other Federal agencies.

All four of our agencies maintain staffs of survey methodologists outside the program offices. Program offices may include methodologists in survey design or redesign projects, and the methodologists' role becomes that of consultant. Methodologists in all four agencies then work collaboratively with the internal sponsors, who are program area subject matter experts. EIA's methodologists participate as consultants throughout most of the QD&T process, including the identification of survey requirements. In addition, a few BLS methodologists are assigned to program offices, where they work directly on survey project teams throughout QD&T.

At the other agencies, the internal sponsors take the lead to determine survey requirements and to develop specific measurements, and methodologists are less involved. However, methodologists at all four agencies participate actively in subsequent QD&T steps. Typically subject area staff draft the initial set of questions, although methodologists may do so at GAO. Subsequently, development of the questionnaire becomes an iterative process between survey methodologists and the sponsoring internal staff. As content specialists, internal staff review draft forms to ensure that content issues are correctly addressed, while survey methodologists review the draft forms for wording, navigation, and layout problems.

Methodologists coordinate questionnaire development and testing activities. At EIA, this includes organizing and conducting focus groups with stakeholders and leading brainstorming sessions with internal program staff to determine requirements. At all of our agencies, methodologists plan and conduct cognitive testing and other pretesting activities. Sponsoring staff often participate in cognitive interviews as observers; at GAO this is required. Following testing, the methodologists summarize results, make recommendations for revisions, and negotiate mutually acceptable resolutions with survey program staffs to address those recommendations.

All four agencies also consult stakeholders and outside subject experts, especially in identifying the requirements for a survey. The U.S. Office of Management and Budget (OMB) requires such consultations for the Census Bureau, BLS, and EIA, as part of the approval process under the Paperwork Reduction Act. Since outside groups can challenge proposed surveys, it is to the agencies' advantage to anticipate the OMB process and involve the outside groups at an early stage in the planning.

None of our agencies have standardized approaches for the outside consultation activity. Instead, a variety of techniques are used, representing a continuum from the more extensive and formal methods used by the Census Bureau and BLS, to others used by EIA and GAO that are less formal. Typically these activities are conducted by the program areas, with little or no involvement by the agencies' methodologists. EIA is an exception; here the methodologists participate as well.

The Census Bureau solicits feedback from trade associations through formal mail-outs of draft or sample forms, particularly for the quinquennial economic census. The agency asks for information about additions and deletions to the form, additional questions that should be asked, and the terminology used. The Census Bureau has recently developed the capability to receive formal comments through the Internet, where individuals can see current and new questions and then provide comments electronically. Trade associations and other stakeholders, including key respondents, are consulted at meetings, by telephone, and through correspondence. In addition, the Census Bureau convenes a formal panel, the Census Advisory Committee of Professional Associations, gathering representatives from academia and private industry to discuss a wide range of issues related to methodology, data collection and data quality. Finally, to ensure that

appropriate data are collected to support National Accounts, the Census Bureau remains in close contact with its primary stakeholder, the Bureau of Economic Analysis.

BLS has used a formal consultation approach with its cooperating State agencies. Prior to one questionnaire redesign, State agency staff completed a questionnaire concerning details of items on the survey form, providing feedback on contents as well as specifications for non-content survey requirements. BLS holds annual meetings with its State partners to engage in discussions regarding program needs and survey procedures. To evaluate specific questions or survey procedures, BLS has also consulted outside experts. For example, when one BLS program office was unable to resolve conceptual issues internally, the office consulted several prominent economists and used their guidance in making final decisions.

On the other hand, in less formal approaches, EIA and GAO generally select and contact stakeholders and agencies with a potential interest in the survey topic to obtain their perspectives on particular issues. The information may be obtained in-person, by phone or by mail. These agencies attempt to identify needs for data among the stakeholder community. Both GAO and EIA discuss the kinds of questions the respondents or industry may or may not be able to answer, and whether or not they have the data. EIA uses this method to develop a model of the subject business or industry and the way it operates, and subsequently assesses its validity during meetings with industry representatives.

2. Conducting Focus Groups

The Method. Focus groups are discussions with a small number of individuals in which the exchange of ideas is guided by a moderator and based on a planned discussion guide or protocol. The intent of a focus group is to collect qualitative information about the participants' views on a given topic or issue. Depending on the purpose of the focus group, participants can be potential respondents, data users, interviewers, or others with some interest in or connection to the survey.

Implementation at our Agencies. BLS, EIA, and GAO have used focus groups for establishment surveys at different phases of QD&T, for purposes as varied as identifying data requirements, assessing measurement issues associated with potential questionnaire concepts, obtaining interviewer perspectives on problematic questions, or encouraging industry participation in an upcoming data collection activity.

Both BLS and EIA have conducted focus groups to aid in identifying data requirements. EIA has conducted focus groups with data users ranging from Federal agencies and Congressional staff to academics, consultants, and the energy trade press, as well as with data providers (e.g., Carlson et al., 1993). BLS has also used focus groups to define or refine survey concepts. For example, on the Employment Turnover and Job Openings pilot survey, focus groups with personnel specialists considered the meaning and measurement properties of a number of concepts such as job separations and new hires (Phipps et al, 1993), while similar groups explored establishment training programs during development of the Survey of Employer Provided Training (Kydoniefs and Horrigan, 1993).

At EIA, the reason for conducting focus groups has evolved over time. The agency discovered that focus groups investigating survey requirements provided the same information as its own subject matter experts. Thus, EIA has come to view focus groups as another means to gain support from data users and data providers for new or redesigned data collections (Freedman and Rutchik, 2002).

GAO has taken a somewhat different approach to focus groups. In addition to convening them for the purpose of identifying survey requirements, GAO has conducted “group cognitive interviews” to pretest instruments with respondents. These special-purpose focus groups have taken one of two formats. In one case, a designated person completed the questionnaire for the group, followed by a discussion of each item among group members. Alternatively, each group member filled in a separate questionnaire, after which they compared responses. In either case, GAO staff observed the discussion and conducted a group debriefing session.

3. Conducting Pre-survey design visits

The Method. It is common survey practice to conduct preliminary research with respondents or potential respondents prior to actually designing questionnaires. For establishment surveys, this research usually involves site visits to provide background information and to assess the feasibility of actually obtaining data that the agencies propose to collect. These visits can occur at various points in the questionnaire development process. During these visits, agencies ask respondents about the existence or availability of specific types of data, along with recordkeeping practices that affect the ease of retrieving the data and the respondent’s willingness to supply that information in the requested format. Agencies may also discuss data sensitivity and timing issues, such as potential release dates. In some cases, if respondents are not able to provide the requested data, discussions may focus on alternative sources of those data or on estimation strategies and the quality of these estimates. Site visits may also cover terminology for use on the questionnaires, respondent understanding of concepts, or the basic structure of an industry. Depending on where they fall in the development cycle, pre-survey design visits can be based on loosely structured discussion guides, topic lists, or semi-structured questionnaires with open-ended questions and detailed probes.

Implementation at our Agencies. All four agencies meet with respondents at some point in the questionnaire development process, to explore general or specific issues associated with a new survey or with a questionnaire that is being redesigned. These meetings have a variety of names: pre-survey design visits (EIA), company visits (Census Bureau), “cognitive interviews” (BLS), and personal visits (BLS and GAO). The goal of most visits is to obtain enough information to construct a questionnaire, but they may also be used for other purposes.

The majority of pre-survey design interactions are face-to-face and take place at the respondent’s business location. They may include subject matter personnel as well as survey methodologists. BLS and EIA primarily conduct in-person site visits. The Census Bureau and GAO make both personal visits and contacts by telephone. The mode decision depends primarily on available resources, as well as on the importance of the survey and whether the questionnaire is new or a revision. For example, internal subject matter specialists at the Census Bureau often turn to known business contacts when forms are being redesigned, using telephone calls as an inexpensive means of getting respondent feedback about proposed new questions or questionnaires. The Census Bureau has also mailed proposed new questions for the next iteration of an ongoing survey to business respondents and followed up with personal visits or by telephone to review the proposed questions.

Our agencies tend to conduct pre-survey design visits for new surveys after identifying initial requirements for the survey, but before preparing draft questionnaires. However, visits have also taken place at later stages in the development process. Regardless of when they occur, at BLS and the Census Bureau, the visits are guided by either a loosely-structured topic list or by a semi-structured interview protocol. The list or protocol may cover respondent comprehension

of concepts or terminology as well as detailed probes about the survey topic. EIA uses a standard protocol that emphasizes data availability and recordkeeping, but also addresses the structure of an industry and whether previously available information can still be obtained or whether it must be acquired from another source.

4. Integrating Content and Noncontent Elements of the Questionnaire

The Method. Some aspects of data collection are independent of survey content. In order to take advantage of new technologies for data collection, certain non-content aspects of the questionnaire need to be considered during forms development. Examples of non-content requirements that impact questionnaires include:

- Postal regulations dictating the position and fonts for preprinted mailing (name and address) information, particularly for materials placed in window envelopes
- Automated mail handling systems requiring specific marginal markings or other identifiers for folding or stuffing
- Information needed for processing, such as bar codes, scanner position markings, and Optical Character Recognition (OCR) or key-entry markings
- Information required by OMB, such as approval number, expiration date, a “purpose and use” statement, and a “time of completion” statement
- Key entry codes for questionnaire coders and respondents who import data into electronic instruments
- Length restrictions or hardware-induced requirements such as printing only on one side of a page
- Aspects unique to electronic data collection such as the user interface, security, data submission, and importing features

These non-content requirements can influence question sequence, question wording, the number of questions asked, and the layout of self-administered questionnaires. Such requirements may lead to reconsideration of questionnaire content.

Implementation at our Agencies. All of our agencies have non-content elements that must be addressed, either at the time data requirements are identified or when data collection instruments are being drafted. Survey methodologists within the agencies work with subject matter and systems staff to ensure that respondent needs are addressed.

BLS, Census Bureau, and EIA print information stored in databases onto their forms for respondents to verify or correct. These agencies also have special bar codes that need to be positioned in specific locations on the survey form for imaging or OCR purposes. On BLS forms, the location of this bar code can vary even for a single survey, if the data are collected by individual states. Within the Census Bureau, offices in charge of form imaging and OCR have standards that affect questionnaire layout.

Electronic instruments also have non-content requirements. For example, the Census Bureau created a style guide for electronic census forms listing non-content requirements for the software developer (Harley et al, 2001).

5. “Borrowing” Questions From Other Questionnaires

The Method. The agencies evaluate questions used previously and “borrow” questions from existing or old surveys. This is particularly important in redesigning questionnaires that support long-standing data series, but is also helpful in creating new questionnaires.

Implementation at our Agencies. All four agencies “borrow” questions and questionnaire formats from existing surveys for new or redesigned agency surveys during the requirements and instrument drafting steps. The Census Bureau has done this regularly while BLS and EIA have drawn questions from past surveys on a more ad hoc basis. GAO has “borrowed” questions from other existing reputable questionnaires or replicated questions from earlier surveys, whenever feasible.

At the Census Bureau, research into previously used questions has been one of the first steps in constructing new surveys or adding new questions to ongoing surveys. For example, during development of a new survey on computer security, as well as the redesign of the Survey of Business Owners, extensive research was done into questions currently used in other economic or demographic surveys within the Census Bureau and other government agencies. During the redesign of annual surveys for the service sector, staff members reviewed other surveys for formatting and layout guidance in addition to question wording.

BLS has borrowed questions and used them in another survey when the goal was to obtain the identical concept from each data collection. As an example, the Job Openings and Labor Turnover (JOLTS) survey uses the same employment questions as the Current Employment Statistics (CES) survey. Questionnaire formats/layouts have also been drawn from other questionnaires. In addition to question wording, the entire JOLTS questionnaire was patterned on CES. Questions may also be borrowed from prior questionnaires if data issues arise that are similar to those in prior research.

Similarly, the American Gas Association (AGA) Underground Weekly Storage Survey recently became an EIA survey. EIA renamed it and made only few substantive changes to the questionnaire. However, the format was changed to meet EIA design requirements.

6. Conducting methodological expert review

The Method. Expert reviews solicit feedback on the content of questionnaires from in-house and external subject matter experts. In addition, agency staffs seek out input on the design aspects of the questionnaire from experts in questionnaire design and survey methodology.

In general, expert review is a process by which experienced questionnaire designers evaluate questionnaires for potential problems with question wording or layout, navigation, or other problems that might interfere with comprehension and response. These reviews focus on question wording, layout of the form, question navigation, and skip patterns. Results of the expert review help to identify those areas of a questionnaire that impose the greatest cognitive, task, and organizational demands. This helps point out potential sources of measurement error.

Implementation at our Agencies. All four agencies use in-house methodological experts to review draft or final versions of questionnaires. The agencies have done this during the initial design or redesign of self-administered forms. The Census Bureau and BLS conduct these reviews at the request of subject matter or program areas, while GAO requires another survey methodologist to review a questionnaire prior to data collection. At EIA, methodologists have reviewed questionnaires prior to the pretest phase.

Census Bureau methodologists typically provide their sponsors or survey program managers with written reports that contain suggested revisions. They may also convene

discussions between survey methodologists and clients. GAO's reviews are less formal, as the agency accepts written comments on the questionnaire. In addition to in-house reviews, both GAO and EIA have used outside consultants to review their questionnaires. For example, GAO contracted with survey methodologists from academic institutions to review a questionnaire about the impact of a specific immigration law. EIA asked academic methodological experts to review questionnaires related to energy consumption and natural gas price and volume.

7. Cognitive Pretest Interviewing

The Method. Cognitive interviews are in-depth interviews, usually conducted one-on-one, with a small number of members of the survey target population. They are designed to acquire information and insight about the "thinking process" used by respondents when answering survey questions. The agencies give the label "cognitive interviews" to a number of different types of interviews that take place at different phases of the QD&T process. These include think-aloud interviews, debriefing interviews, exploratory interviews, and group interviews.

Depending on the QD&T phase, the interviews may involve watching a respondent complete a questionnaire, asking a respondent how they prepared a questionnaire that was mailed in advance, or asking respondents how they submitted data in a long-term pilot study. Cognitive interviews may also include questions about a respondent's basic understanding of survey concepts, about the availability of data, or about how closely a respondent's business records conform to survey concepts. They may also include topics that are covered in pre-survey design visits.

Implementation at our Agencies. All four agencies conduct cognitive pretest interviews of draft survey forms with actual business respondents. The interview method depends more on the needs of the study than on the agency that is conducting it, and there is no one set way for conducting a cognitive interview in any of our agencies. Interviews can vary by the number and types of participants involved, the type of interview protocol used, and whether or not they are audio taped.

Interviews generally take place at business locations, allowing respondents access to records and encouraging discussion of the data retrieval process. The number of survey methodologists and agency subject matter staff participating in the interviews may vary. For the Census Bureau and BLS, generally one or two observers accompany the researcher on the interviews. The observers include survey sponsors, subject matter analysts, or other survey methodologists. While EIA participants can include survey methodologists and subject matter specialists, GAO actually *requires* participation by one methodologist and at least one subject matter analyst. The rationale for the latter requirement is that if a respondent asks for information about the subject matter, someone with that knowledge is readily available to answer these questions. Further, having subject matter specialists attend these interviews can facilitate making future corrections and changes to the questionnaire.

The Census Bureau and BLS develop an interview protocol to use as a guide for the interviews. Depending on the objectives of the research, the protocol may contain different combinations of think-aloud methods, concurrent questions, debriefing questions, or hypothetical probes. The researcher may also choose to use vignettes for part or the entire interview. With the permission of the respondent, cognitive interviews are audio recorded to facilitate accurate summarization.

The EIA process for on-site interviews has three components. The first component includes a formal 'think aloud' interview with structured probes, in which respondents are asked

to go through the questionnaire and describe how they would complete the survey form. The second component includes a discussion with the respondent about the establishment's record keeping practices. For example, EIA investigates whether operational definitions of terms apply to the respondents, and whether billing cycles are compatible with EIA's need for calendar month data. In the third component, survey methodologists and subject matter specialists answer respondent questions about the survey (Rutchik and Freedman, 2002).

GAO almost always uses a retrospective protocol and respondent debriefing technique. In this process, the respondent completes a questionnaire while GAO staff observe. The survey methodologist makes notes of respondents making any erasures, cross-outs or pauses, exhibiting any body language, making unexpected page turnings or making utterances. The methodologist also times item completion. The debriefing session which follows includes an item-by-item review of the questionnaire. The researcher asks non-directive probes about the meaning of terms, rephrasing of questions, and the like. At the end of the session, GAO also asks general questions about potentially sensitive items, the length of the questionnaire, and items to add or take out (Featherston and Moy, 1988: 16-21). GAO does not audiotape these interviews, instead relying on detailed notes. GAO also conducts "group cognitive interviews" as described in Section III.A.2.

In addition to the above techniques, BLS, GAO, and sometimes EIA conduct cognitive interviews over the telephone. For example, BLS has conducted these by asking respondents to use a concurrent or retrospective 'think-aloud' protocol, using faxed documents as the basis for the discussion.

8. Testing Using Vignettes

The Method. In establishment survey QD&T, vignettes are either descriptions of fictitious situations or mock records resembling the internal documents of a business. Respondents are asked to interpret the situation accordingly, or they are asked to consult the mock records to complete the form. Vignettes commonly supplement traditional think-aloud cognitive interviewing, but they can also be used when pretesting self-administered questionnaires.

Implementation at our Agencies. BLS and the Census Bureau have used vignettes to test paper and electronic forms. For self-administered paper questionnaires, vignettes were used to test alternative form layouts. Respondents completed questionnaires using mock records (Stettler et al., 2000). For electronic forms, vignettes were used to test the layout, functionality, and navigation of the software, and respondents were asked to use the software to complete specified tasks or scenarios. The implementation of vignettes differs between the agencies. The Census Bureau has incorporated vignettes into concurrent think-aloud cognitive interviews. At BLS, respondents completed a form using vignettes and then answered a self-administered debriefing questionnaire.

The Census Bureau has also used vignettes to test survey concepts (Morrison et al., 2002). In testing questions on employee leasing, methodologists created vignettes about various alternative employment arrangements. Researchers explored business respondents' understanding of this newer term and observed reactions to it. Another recent use of vignettes tested the alternative wording of proposed edits for an electronic instrument (Anderson et al., 2001). A paper prototype of an on-line mortgage application was pre-filled with erroneous data. Respondents, who were internal Census Bureau employees, interpreted the error messages and corrected the data. This allowed researchers to identify problems with the alternative edit message wording or presentation and to recommend changes.

9. Questionnaire Pretesting

The Method. Questionnaire pretesting refers to the administration of a draft questionnaire to a small sample of respondents who meet all criteria for regular data collection, for the purpose of identifying problems with the questionnaire. Operationally, pretesting usually involves conducting telephone interviews and often entails paper-and-pencil data collection prior to programming a questionnaire for a CATI system.

Questionnaire pretesting differs from pilot testing in two ways. First, the number of cases is generally small. Second, the pretest results are reviewed immediately, with very short turnaround on revisions. To facilitate that review, interviews may be monitored by survey methodologists or interviewing supervisors. As soon as the pretest is over, the staff monitoring the calls evaluates the results and makes needed revisions to the questionnaire.

Implementation at our Agencies. BLS is the only agency among the four that has conducted this type of questionnaire pretesting. In the early stages of a pilot study carried out by a contractor for BLS, interviewers conducted several rounds of paper and pencil pretesting for a telephone script. Each test round resulted in questionnaire modifications, after which the script was programmed for CATI administration (Goldenberg et al., 1997). In another case, State staffs from two Federal-State cooperative programs conducted a telephone pretest of a respondent debriefing survey comparing responses to the two programs. The program staff used the pretest results to revise the questionnaire prior to going into the field for formal collection. In the situations described here, interview scripts were based on pre-survey design visits that included “cognitive” components such as respondent understanding of concepts. These scripts drew heavily on the knowledge and experience of subject matter staff. Neither of the questionnaires in these examples had been pretested with cognitive think-aloud or debriefing interviews.

10. Functionality and Usability testing of Electronic Instruments

The Method. Functionality and usability testing are methods for testing electronic data collection instruments. Electronic instruments include screen-based instruments, such as those used in Internet surveys or computer automated telephone interview systems, and non-screen-based instruments such as touchtone data entry or hand-held data entry systems. Although functionality and usability testing are separate testing methods, they are often done concurrently. As more agencies develop electronic interfaces for their surveys, these testing methods will become an increasingly important element in QDET.

The goal of *functionality testing* is to uncover any unknown software problems that need to be addressed prior to release of the instrument to respondents. This process typically consists of usability experts, subject matter experts, or systems staff testing the computer functions of the electronic instrument to ensure that the application operates as intended. These tests might include verifying that the respondent can access the instrument, the arrow keys move from question to question, the respondent can exit the document, and the scroll bar can be seen or moved. Reviewers try various combinations and permutations of steps in an attempt to “break” the application.

Usability testing is often conducted to review newly developed or existing electronic instruments. The main purpose of usability testing is to assess the interaction between the computer and the respondent, as well as any effects the user interface may have on the content of the questionnaire. This includes ensuring that skip patterns are working appropriately, the navigation buttons move the questionnaire to the appropriate spot, or questions are readable on

the screen. The examination of content issues may not be the focus of a usability test but often occurs regardless.

The process sometimes begins at the questionnaire design stage where human-computer interaction experts conduct “heuristic reviews” of the usability of the electronic forms interface in lieu of or before usability testing. Heuristics are a set of established usability guidelines. The purpose of the heuristic review is to pinpoint usability violations by comparing the application to these rules and providing recommendations that can allow programmers to modify an application before usability testing. Some usability testing is conducted in a laboratory that is equipped with cameras focused on the participant’s face, the computer screen, the keyboard, and sometimes the mouse.

Implementation at the Agencies. Functionality testing is routinely conducted by all of our agencies prior to the distribution of an electronic form to respondents. The methods or strategies vary among the four agencies. For example, EIA has used external experts to conduct functionality testing primarily on web-based instruments, while Census Bureau, BLS and GAO staff primarily use internal experts. BLS has also conducted functionality testing for non-screen based instruments. For example, for a touchtone data collection survey, BLS program office staff tested the instrument by having staff members make simultaneous call-ins prior to going “live” with the survey. All of these efforts are agency attempts to “stress” the electronic data collection system to ensure the smooth operation of the data collection effort.

All four agencies test usability of electronic instruments to assess computer functionality, navigation, and layout, and conduct these tests at various stages of electronic instrument development. BLS, the Census Bureau, and GAO have conducted expert or heuristic reviews of the usability of the electronic forms interface. The Census Bureau has drawn on usability experts, and sometimes questionnaire design experts, to conduct this type of review. BLS and GAO have included personnel from program offices and internal subject matter experts as reviewers. Reviewers have been asked to examine the electronic forms and to point out problems with navigation, layout, functionality, instructions, and edits that could be changed before the instrument was tested with respondents. This activity has typically occurred before formal usability testing but has sometimes coincided with usability testing.

Usability testing can be conducted in the laboratory or on-site at a business location. The Census Bureau and GAO have conducted usability tests on-site at respondents’ business locations. Census Bureau researchers have tried to mimic a laboratory setting by videotaping the test, with the permission of the respondent. During the test, respondents were shown certain screens and asked questions about them or they were asked to perform various functions based on a predetermined scenario (i.e., vignette). They were asked to think aloud during this entire process while researchers observed their actions, keystrokes, etc. The GAO researchers, on the other hand, observed their respondents completing a draft version of the questionnaire on the Internet. Researchers from both agencies asked probing questions during the interview. The interviews were then followed up with some debriefing questions focusing on overall impressions of the electronic form. Respondents were asked to discuss specific pieces of the form that they especially liked or disliked.

BLS conducts its testing at an in-house usability laboratory, which allows close observation of the participant. Respondents may be asked to complete a short self-administered questionnaire following the test session. In addition to in-lab usability testing, BLS has also conducted usability testing on handheld pen-tablet computers in a group setting. Users, who were in this case data collection personnel, were asked to complete the data entry activities they did on

their job, and later to enter data based on vignettes. Users worked in pairs to perform each task, while the researcher observed their efforts and noted difficulties. The users then participated in a group debriefing session, similar to a focus group, to address problems and issues in actual use of the software.

Usability testing was conducted by the Census Bureau in its usability laboratory to support development of a style guide for electronic economic census forms (Anderson et al., 2001). A style guide is a list of rules encompassing areas such as screen layout, editing, graphics, and functionality that the software developer must apply when developing the final electronic forms. A team consisting of survey methodologists, usability experts, electronic form experts, subject matter specialists and developers constructed the rules according to current usability principles, past usability tests done on similar electronic forms, and tests conducted solely for this style guide (Murphy et al., 2001).

B. Techniques and Approaches used in Questionnaire Evaluation

As described above, the distinction between QD&T and QE is that QD&T takes place prior to collecting data with the questionnaire, while QE follows it. One of the goals of QE for an ongoing survey is to provide information that can be used in the redesign of the questionnaire. For new questionnaires, pilot tests of the survey instrument and assessments of response issues and data quality from fielded questionnaires provide sources of QE data. Other important sources of evaluation data are the experiences of individuals who have direct contact with respondents, such as data collectors or members of a survey staff. Respondent debriefings and reinterviews are other useful evaluation tools.

1. Conducting Pilot Tests

The Method. Formal pilot tests or field studies, which involve conducting the survey with a subset of the target population, are an important tool for evaluating data collected by new or redesigned questionnaires. In many cases, the primary goal of a pilot test is to evaluate a questionnaire or data collection approach, and the test permits researchers to observe its performance in the context of other survey operations. Pilot tests may seek to assess the questionnaire's effectiveness in collecting the desired data, to examine data quality in terms of item nonresponse and edit failures, or to review the potential for, frequency of, and types of errors made by respondents. Other goals include evaluating respondents' understanding of questions and definitions, investigating response issues, refining the questionnaire, and testing the feasibility of the data collection plan. Depending on the evaluation goals, sample sizes and designs for pilot studies may range from small convenience samples to stratified random samples of several thousand businesses, and may represent an underlying population or be limited to businesses having characteristics most pertinent to the redesigned questions.

Pilot tests may be full-scale "dress rehearsals," that is, operations tests designed to test all of the steps and systems in the data collection process, including estimation. Alternatively, if the primary goals are to assess the effectiveness of data collection procedures and the performance of the questionnaire, the test may not include estimation or preparation of detailed tabulations.

Implementation at our Agencies. Two of the agencies, the Census Bureau and BLS, have conducted pilot tests prior to full-scale data collection, both for new surveys and for surveys with redesigned questionnaires. These agencies have used pilots in two ways: to test an entire data collection process and to carry out a detailed assessment of the pilot questionnaire results. In some cases, these agencies have incorporated split-sample designs into the studies, allowing

comparisons of alternative questionnaires, alternative collection procedures, or data obtained using different concepts. (See Section III.B.3.) Many of these pilot tests followed rigorous iterative approaches to questionnaire development, including several rounds of cognitive pretesting and consultation with stakeholders and sponsors.

There are a number of differences between the two agencies in their use of pilot testing. First, Census Bureau studies tend to be larger than their BLS counterparts. They can involve many thousands of respondents, and a 500-unit pilot test is considered “small.” In contrast, BLS pilot studies involve a few hundred to at most two to three thousand respondents; when the study is an ongoing one with monthly data collection, the number of units may be in the low hundreds. In addition, the Census Bureau studies test mail-out, mail-back self-administered questionnaires, distributed through its National Processing Center. Many BLS pilot tests have been of questionnaires distributed by mail, with data collected by telephone interview for some period of months afterward, and BLS has generally contracted out its pilot test data collection. A third difference is that the Census Bureau tends to pilot test questionnaires that are very close to final, while BLS is more likely to conduct tests somewhat earlier in the questionnaire development process and to focus more on feasibility issues or specific data collection issues, with less emphasis on response error.

There are also similarities between the agencies’ uses of pilot testing. Both the Census Bureau and BLS have used pilot tests to study the data collected from test questionnaires, looking at response distributions, skip pattern errors, and item nonresponse. (BLS would call this type of test a “pretest.”) Census Bureau researchers have evaluated patterns of answers for appropriateness or logical consistency and for the number of “fatal” errors that rendered a response unusable. In one case, a Census Bureau pilot study tested a redesigned questionnaire, enabling subject area staff to become familiar with response patterns before launching the full-scale survey.

Where pilot tests have been designed to compare specific questions, both the Census Bureau and BLS have embedded split sample designs into pilot tests. At the Census Bureau, these designs have been used to compare new questions with those previously used. BLS has more often used split samples to evaluate solutions for procedural problems or to determine whether certain types of data might be more readily reported than others (section III.B.3).

Because of the evaluative nature of pilot tests, the Census Bureau has typically included follow-up evaluations as part of its data collection process. Self-administered evaluation forms have been enclosed with the mailed survey questionnaires, asking the respondent to estimate the amount of time needed to complete the questionnaire, the extent to which records were consulted, and which questions were problematic. If appropriate, such an evaluation has included questions to further address issues that arose during cognitive testing. Sometimes part of the sample has been debriefed by telephone instead, allowing researchers to probe for more detail. A BLS pilot followed a similar approach, with a page of debriefing questions at the end of a mailed questionnaire. Interviewers called back one group of respondents and asked detailed questions about data on the form and about questions that were not answered. In another BLS pilot study, interviewers made monthly data collection calls for a year. Each month respondents were also asked supplementary questions that helped the survey staff better understand the issues and problems associated with the reported data (Mueller and Phillips, 2000).

2. Analyzing Collected Data

The Method. Analysis of item nonresponse rates, imputation rates, or edit failures from the collected data can provide useful information about how well the questionnaire works. In some surveys, high item nonresponse indicates that data are not routinely available in records. This may lead to the removal of questions or to redefining response categories. In still other surveys where a series of related figures are collected, item missing data, along with inconsistent data, are detected by edit failures and may be subject to imputation. In addition, the quality of collected data may also be evaluated by making comparisons with data from other sources. Depending on the subject matter context, item nonresponse rates, imputation rates, and/or edit failure rates, as well as other signs of problem data, may be used to indicate questions warranting further investigation and/or evaluation by other methods.

Implementation at our Agencies. All four agencies analyze their collected data as part of questionnaire evaluation. Whether this analysis feeds back into QD&T depends on the agency. EIA routinely compares its petroleum data with data from other sources, and compares data from its energy consumption surveys to data from surveys of energy suppliers. These comparisons sometimes result in changes to survey forms, because of discrepancies observed between two surveys where data should be comparable.

Both the Census Bureau and BLS routinely monitor survey imputation rates and edit failures, and reconcile reported data for some surveys with data from other sources. For example, the aggregation of data reported on monthly Census Bureau surveys is benchmarked against annual survey results. At the company level, reported data may be compared with figures found in publicly available sources, such as annual reports. However, the main purpose of most of these analyses has been to assess data quality or data processing methods, not to improve questionnaires.

The BLS Annual Refiling Survey asks respondents to verify or change the industry description printed on the form for their establishments. The data system identifies cases where respondents say the description is not correct, but additional respondent-supplied information indicates that the establishment was coded correctly. High levels of “incorrect-correct” cases in an industry suggest that the description is not clear and needs to be rewritten, thus feeding directly back to questionnaire content. Similarly, high levels of missing physical location addresses have been traced, at least in part, to instructions that were easily overlooked and that were often ignored when the physical location and mailing addresses were the same.

GAO also reviews edit failures, item nonresponse, and internal data inconsistencies, and may compare the data from a survey to other data sources. However, these data do not normally affect questionnaire redesign, since many GAO surveys are one-time only. On the other hand, when GAO borrows questions from previously used questionnaires, survey methodologists may take into account information such as edit failures and item nonresponse.

3. Split sample testing

The Method. Split sample testing incorporates an experimental design into data collection, allowing researchers to empirically compare one or more alternative questionnaires or questions. In a split sample design (also known as “split ballot” or “split panel” experiments), each alternative questionnaire, or “treatment,” is randomly assigned to a sample unit. As a controlled experiment, results from each treatment can be compared and the “better” alternative assessed. Split sample tests are often conducted in conjunction with pilot/field tests but may also be incorporated in ongoing or recurring surveys. The latter are important to establishment surveys,

particularly those with the purpose of measuring change. Split sample testing then allows comparison of results from “new” versions of questions or new data collection procedures with “old” questions or procedures, so researchers may determine the effect on statistical time series of changes in methodology (Federal Committee on Statistical Methodology, 1983; Gower, 1994).

Implementation at our Agencies. Among our four agencies, only the Census Bureau and BLS have conducted split sample tests during data collection, and even then, they have not been conducted frequently. Split sample designs are typically embedded into a pilot test rather than into operational data collection. For example, the Census Bureau recently conducted a split sample experiment within a pilot test of a redesigned survey collecting characteristics of business owners. The split sample allowed statistical comparison of the previous form with the redesigned form in terms of reporting differences and response rate differences. At BLS, split sample tests have compared the collectibility of earnings data using two different concepts of earnings (Goldenberg et al., 2000), and the use of alternative sets of instructions to reduce calendar-related fluctuations in reported employment and earnings (McConnell et al., 2001). Both of these studies have been part of pilot tests.

In addition, the Census Bureau and BLS have used split samples to assess aspects of the survey process other than questionnaire evaluation. For example, the Census Bureau included a split sample in an operational survey of manufacturing establishments to study the effectiveness of alternative strategies offering a Web reporting option (Dodds, 2001). Similarly, a BLS experiment compared alternative mailing materials for the Annual Refiling Survey to see if one approach was more effective than others in encouraging sample members to respond by touchtone telephone rather than mail. Incorporating a split sample design into operational data collection appears to be rare, perhaps because doing so involves a separate processing stream.

4. Debriefing Survey Staff and Interviewers

The Method. Past knowledge from the experience of staff is invaluable when revising questionnaires or developing new ones. Because of their direct interaction with respondents, survey staff involved in day-to-day survey operations are well-positioned to identify problems or issues with data collection instruments. Survey personnel include program staff, analysts, and interviewers, including CATI interviewers when they serve as the primary data collectors. These staff members can offer insights into sources of confusion and problems with data submissions or the survey instrument because of their direct interaction with respondents. Debriefing the staff can be as formal as a focus group or as informal as observations presented during a regular staff meeting. These staff debriefings provide input to the agencies as they review and modify survey instruments for future data collection cycles or for related surveys, thereby linking the questionnaire evaluation phase back to the questionnaire development and testing phase.

Implementation at our Agencies. BLS, the Census Bureau, and EIA systematically or informally collect and evaluate data from their staffs as they review respondent behavior and the overall data collection experience. The approach taken by each agency varies.

EIA evaluates an existing survey through organized “brainstorming” sessions with survey staff. Facilitator-led discussions involve survey managers, individuals who screen returned questionnaires, data editors, staff who telephone respondents to resolve data problems, and personnel who answer calls from respondents (Freedman and Rutchik, 2002). Similarly, the Census Bureau technical staff who have direct contact with respondents are encouraged to maintain a log of questionnaire design problems and issues.

At BLS, state employees who are responsible for data collection under the Federal-State Cooperative Program comment informally about specific issues. During one survey redesign, the BLS program office sent the states a questionnaire asking about certain issues regarding the form. This self-administered debriefing contributed to significant changes in form design and layout.

For CATI data collections, BLS has conducted occasional debriefings using focus groups or individual meetings with interviewers. While these debriefings often address procedural aspects of the survey (e.g., how interviewers encourage survey participation), they also help to pinpoint problem areas in questionnaires and other aspects of data collection (Fisher, 2002).

Some self-administered paper and electronic surveys at the Census Bureau include a CATI nonresponse follow-up component. When these surveys were redesigned, supervisors monitored the follow-up interviews to see how respondents handled the new questions. During questionnaire redesign interviewers have sometimes been asked for feedback.

5. Debriefing Respondents

The Method. A respondent debriefing is a contact with a survey respondent following that person's participation in a survey or data collection activity. The purpose of the debriefing is to evaluate the reported data, the data collection instrument, or both. More specifically, a respondent debriefing provides information that allows survey personnel to assess the quality of the reported data, to learn about the respondents' interpretation of the questions and terminology used within the survey, and to identify the data sources that the respondent consulted to answer the survey questions. These debriefings may be administered in person or by telephone, or they may be self-administered. They often include detailed questions about specific characteristics of the data reported. Debriefings can be incorporated into any data collection environment, and some form of debriefing is often included as an assessment tool in a pilot study.

One form of respondent debriefing is called a Response Analysis Survey (RAS). In a RAS, researchers make telephone contacts with respondents who have returned completed mail questionnaires or who have previously been interviewed (usually by telephone) for a survey. Using a scientifically-representative sample of the survey's participants, a RAS is based on a structured questionnaire and generates quantitative data. A RAS may include some qualitative elements, such as respondents' interpretation of specific terminology. However, one difference between a RAS and most cognitive pretest interviews is that the RAS interview is administered to a subsample of respondents from the original survey. As a consequence, results can be generalized to the larger study population. RAS studies are especially useful for ongoing surveys, where their results can help inform proposed questionnaire revisions.

Implementation at our Agencies. Both BLS and the Census Bureau have conducted systematic debriefings with respondents. Due to resource constraints, most Census Bureau respondent debriefings consist of self-administered evaluation forms enclosed with the survey questionnaire. Questions are typically open ended, asking respondents to identify problem items or to indicate whether records were used. Sometimes specific questions have been included to address issues that were raised during cognitive pretesting. BLS has also conducted respondent debriefings with short self-administered questionnaires during pilot testing, primarily as a means of assessing response burden and question difficulty (Phipps and Horrigan, 1993).

Both of these agencies have also conducted respondent debriefings by telephone. While a Census Bureau debriefing can be either formal or informal, a debriefing at BLS is generally a formal data collection effort. A Census Bureau debriefing may seek to evaluate the questions or

instructions, while a BLS RAS may also look at unanswered questions and try to determine the reason that those questions were not answered. Some RAS studies also ask questions about the availability of data in specific formats (Goldenberg et al, 1993).

6. Reinterviewing respondents

The Method. Reinterview studies evaluate measurement variance, bias or reliability of reported data by collecting data a second time from survey respondents. Typically the second interview attempts to obtain the “true values” of selected survey variables, with more time and attention given to the response task in the second interview in order to obtain accurate data. Data reported in the original survey may also be reconciled with reinterview responses and the reasons for differences identified, indicating sources of measurement error. In establishment surveys, reinterview methods may require respondents to access records and provide “book values” or to construct answers from data in records when the original responses were estimated values (Biemer and Fecso, 1994). Again, the primary purpose of reinterview studies is to assess data quality; the degree of feedback into questionnaire design or data collection methods varies.

Implementation at our Agencies. Of our four agencies, only GAO and the Census Bureau have conducted reinterview studies. Their experiences and motivations are very different, however. Since GAO typically conducts one-time surveys, it is important to verify survey data when there is concern about the data reliability, especially in situations related to regulatory compliance. In these instances, GAO has indicated in a questionnaire’s introduction that it will select some respondents for a review of responses, and has asked that the respondents retain their supporting materials. In past studies, GAO selected a small percentage of the responding sample and reinterviewed respondents either in-person or by telephone, examining selected data elements. GAO staff have also gone through answers with the respondents during the site visits that are typically conducted as part of the study methodology.

The Census Bureau conducted reinterviews called “content evaluations” in association with the 1977 and 1982 Economic Censuses (Corby, 1984; Corby, 1987). These reinterview studies asked about both the responses for specific data elements and the various components of those responses (e.g., items or people included and excluded, actual data versus estimates, etc.). Based on statistical samples, these studies empirically evaluated the size, direction, and sources of biases in reported data. Though the emphasis was on assessment of data quality, the content evaluations included some recommendations for future questionnaire improvements.

IV. Conclusion

In this paper, we have described our knowledge of and experiences with QDET processes and methodologies used for designing data collection instruments and questionnaires for establishment surveys. Each of the four U.S. government agencies we represent – BLS, the Census Bureau, EIA and GAO – have a specific mission and primary customers for its data. The subjects we survey differ. The domains of BLS and EIA are indicated in their names (labor and energy, respectively), while the Census Bureau and GAO collect data on a wide variety of economic topics. GAO typically conducts one-time surveys, while the other agencies conduct recurring ongoing surveys having weekly, monthly, quarterly or annual frequencies. Although all of the authors operate in a consultancy capacity either within or across program areas, our positions within the organizational structures of our agencies differ.

Probably the greatest and most evident source of our differences lies in our sponsors and stakeholders – who they are, their role and influence in the QDET process, and how our agencies interact with them and they with us. The role of external subject matter experts also varies across our four agencies, where in some cases they exercise influence comparable to that of stakeholders. These differences depend on the position of our agencies within the Federal government and the subject areas we cover. Nevertheless, these sponsors, stakeholders, and external experts, whoever they may be for our agencies, all wield some amount of influence and participate in the QDET process. Our agencies differ to some extent in our methods for consulting with them. On the other hand, most, if not all, establishment surveys in our four agencies have internal sponsors. We survey methodologists interact with these internal sponsor subject matter specialists in a collaborative manner to develop and test data collection instruments.

With little variation, the steps in the QDET process are the same for all four agencies, regardless of the type, topic or frequency of the survey. We all:

- Develop requirements for the questionnaire
- Draft data collection instruments and procedures
- Subject instruments to expert review
- Pretest instruments with respondents
- Revise and review draft instruments
- Examine the quality of data collected using these instruments

We observed great commonality in the QD&T and QE methods we use, their purposes and the points in the process where we used them. This commonality is illustrated in Table 1. The primary exception is GAO, where the predominance of one-time only surveys renders some methods less useful or inappropriate; in other cases, the method is the same but the purpose differs. Methods that appear to be used less pervasively across our four agencies include focus groups, vignettes, questionnaire pretests, pilot/field tests, split sample experiments, respondent debriefings, and reinterview studies.

Despite the commonality in the methods and our implementation of them, we found that our agencies often use different terminology to describe the same or very similar techniques. A parallel difficulty is that our agencies may use a word or phrase to represent different actions and/or different aspects of the questionnaire development process. In fact, terminology was probably the greatest source of differences among our agencies (ironic, since ours is a discipline that literally hangs on every word!). The descriptions provided in this paper focus on the activities rather than on the terminology. A useful outcome of this collaboration has been to work toward a common vocabulary for what are evidently comparable types of work.

Regardless of the differences between our agencies' missions, subject areas and terminology, the similarities in our QDET steps and methods derive at least in part from the fact that we all survey the same target population, that of establishments (businesses, organizations and/or institutions). Factors associated with response processes in this population dictate the methods used and shape their implementations in a common way.

V. References

- Anderson, Amy E., Nichols, Elizabeth and Pressley, Kimberly (2001), "Usability Testing and Cognitive Interviewing to Support Economic Forms Development for the 2002 U.S. Economic Census." *Proceedings of the Statistics Canada Symposium 2001*, Hull, Quebec, October.
- Biemer, Paul P. and Fecso, Ronald, S. (1995), "Evaluating and Controlling Measurement Error in Business Surveys," in Cox, Brenda et al. eds. *Business Survey Methods*. New York: John Wiley & Sons: 257-281.
- Carlson, Lynda T., Preston, John L., and French, Dwight K. (1993), "Using Focus Groups to Identify User Needs and Data Availability." Alexandria, VA: *American Statistical Association, Proceedings of the International Conference on Establishment Surveys*: 300-308.
- Corby, Carol (1984), "Content Evaluation of the 1977 Economic Censuses (DE-2)," *Bureau of the Census Statistical Research Report 84/29*, Washington D.C.: U.S. Census Bureau.
- Corby, Carol (1987), "Content Evaluation of the 1982 Economic Censuses: Petroleum Distributors," in *1982 Economic Censuses and Census of Governments Evaluation Studies*, Washington D.C.: U.S. Census Bureau, pp. 27-50.
- Dodds, Judy (2001), "Response to the 1999 Annual Survey of Manufactures: Computer Network Use Supplement," Internal report. Washington D.C.: U.S. Census Bureau.
- Featherston, Fran and Moy, Luann (1988), "Minimizing Respondent Burden." Paper presented at the American Association for Public Opinion Research, Toronto, Ontario, Canada.
- Federal Committee on Statistical Methodology (1983), *Approaches to Developing Questionnaires*. Statistical Policy Working Paper #10. Washington, D.C.: Office of Management and Budget.
- Fisher, Sylvia K. (2002), "Transitioning CES Respondents to TDE: Results of Focus Groups Conducted with the Atlanta and Kansas City DCC Interviewers." Internal report. Washington, DC: Bureau of Labor Statistics.
- Freedman, Stanley R. and Rutchik, Robert H. (2002), "Information Collection Challenges in Electric Power and Natural Gas." Presented at the Joint Statistical Meetings, New York, NY. Alexandria, VA: *American Statistical Association, Proceedings of the Section on Survey Research Methods*, forthcoming.
- Goldenberg, Karen L., Butani, Shail, and Phipps, Polly A. (1993), "Response Analysis Surveys for Assessing Response Errors in Establishment Surveys." Alexandria, VA: *American Statistical Association, Proceedings of the International Conference on Establishment Surveys*: 290-299.

- Goldenberg, Karen L., Levin, Kerry, Hagerty, Tracey, Shen, Ted, and Cantor, David (1997), "Procedures for reducing measurement errors in establishment surveys." Presented at the American Association for Public Opinion Research, Norfolk, VA. Alexandria, VA: American Statistical Association. *Proceedings of the Section on Survey Research Methods*.
- Goldenberg, Karen, Gomes, Anthony, Manser, Marilyn, and Stewart, Jay (2000), "Collecting All-Employee Earnings Data in the Current Employment Statistics Survey: A comparison of two concepts." Alexandria, VA: American Statistical Association, *Proceedings of the Section on Government Statistics*.
- Gower, A.R. (1994), "Questionnaire Design for Business Surveys." *Survey Methodology* 20 (2): 125-136.
- Harley, M. Diane, Pressley, Kimberly D., and Murphy, Elizabeth D. (2001), "2002 Economic Electronic Style Guide." *Proceedings of the Statistics Canada Symposium 2001*, Hull, Quebec, October 2001.
- Kydoniefs, Leda and Horrigan, Michael (1993), "Summary of Focus Group Research and Revisions to the Survey of Employer Provided Training." Internal report. Washington, D.C.: Bureau of Labor Statistics.
- McConnell, Sheila, Dillender, Rebecca, Logothetti, Ted, and Strifas, Sharon (2001), "Correcting the Calendar-related Fluctuations in the Average Weekly Hours and Average Hourly Earnings Series." Alexandria, VA: American Statistical Association, *Proceedings of the Section on Government Statistics*.
- Morrison, Rebecca L., Stettler, Kristin, and Anderson, Amy E. (2002), "Using Vignettes in Cognitive Research on Establishment Surveys." Paper presented at the International Conference on Questionnaire Development, Evaluation and Testing Methods, Charleston, SC, Nov. 15-17.
- Mueller, Charlotte and Phillips, Mary Anne (2000). "The Genesis of an Establishment Survey: Research and Development for the Job Openings and Labor Turnover Survey at the BLS." Alexandria, VA: American Statistical Association, *Proceedings of the Section on Government Statistics*.
- Murphy, Elizabeth D., Nichols, Elizabeth M., Anderson, Amy E., Harley, M. Diane, and Pressley, Kimberly D. (2001), "Building Usability into Electronic Data-Collection Forms for Economic Censuses and Surveys." *Federal Committee of Statistical Methodology Research Conference*, November, 2001, Arlington, VA.
- Phipps, Polly A., Butani, Shail, and Chun, Young (1993), "Designing Establishment Survey Questionnaires." *BLS Statistical Note 35*. Washington, D.C. Bureau of Labor Statistics.

Phipps, Polly and Michael Horrigan (1993), "Summary Research Report, OMB Clearance Package 1220-0141. Survey of Employer Provided Training Pretesting." Internal report. Washington, DC: Bureau of Labor Statistics.

Rutchik, Robert H. and Freedman, Stanley R. (2002), "Establishments as Respondents: Is Conventional Cognitive Interviewing Enough?" Presented at the International Conference on Questionnaire Development, Evaluation, and Testing Methods, Charleston, SC.

Stettler, Kristin, Morrison, Rebecca and Anderson, Amy E. (2000), "Results of Cognitive Interviews Studying Alternative Formats for Economic Census Forms." Alexandria, VA: *American Statistical Association, Proceedings of the International Conference on Establishment Surveys II*: 1646-1651.