

# Gemini Project Overview

JENNIFER EDGAR and  
ADAM SAFIR

The Consumer Expenditure Survey (CE) program initiated the Gemini Project, a multiyear survey redesign effort, in early 2009. The mission of the project is to promote improved expenditure estimates in the CE by reducing measurement error. During the course of the project, CE program staff will develop, test, evaluate, and (as appropriate) implement survey design changes with the goals of improving overall data quality, increasing the analytic value of the data to users, and supporting greater operational flexibility to respond to changes in the data-collection environment.

The changes being pursued through the Gemini Project will ensure that the CE satisfies its primary purpose: maintaining the integrity of the expenditure weights used in the Consumer Price Index (CPI). Consumer expenditure data supplied by the survey are a critical component in the calculation of the CPI because they are used to estimate weights for the CPI's consumer goods and services classification structure. In the construction of the CPI, CE data serve four distinct functional uses: (1) to estimate annualized expenditures, (2) to estimate monthly expenditure weights, (3) to probabilistically select item categories for pricing, and (4) to allocate expenditure estimates between more broadly defined expenditure categories from other survey sources.<sup>1</sup>

Improved data quality also enhances the usefulness of CE data in meeting the needs of other data users, both public and private. Increasing the flexibility of survey operations allows the program to meet new data-quality challenges in a timely fashion.

This article reviews the background, motivation, and challenges that affect the survey and the survey redesign effort specifically. The article also provides an overview of the project's approach, major activities to date, and an overall timetable, paying particular attention to research topics affecting the redesign, as well as plans to investigate those topics. The article concludes with a summary of project accomplishments and plans for the future.

## Gemini Project background

The overall goal of the Gemini Project is to improve data quality by reducing measurement error, which is the difference between the reality of a respondent's situation—and what the respondent reports to the CE. A major focus will be on underreporting, including error at the individual-item level and at the larger-scale household-budget level. The latter may not matter if the relative shares remain unbiased, but it ultimately is an important feature of underreporting that needs to be addressed. At the same time, a secondary goal will be to maintain or increase response rates; therefore, any expected benefits of survey design changes will be weighed against potential negative effects on response rates.

## Motivation for redesign

All household surveys today face challenges that affect response rates, including busy respondents, confidentiality/privacy concerns, competing surveys, controlled-access residences, and non-English-speaking households. In addition, the CE is faced with particular issues that directly affect the quality of the data collected. These issues, presented in order of importance, include a) evidence of measurement er-

Jennifer Edgar is a research psychologist in the Office of Survey Methods Research, Bureau of Labor Statistics.

Adam Safir is Branch Chief, Research and Program Development, Division of Consumer Expenditure Survey, Bureau of Labor Statistics.

<sup>1</sup> Casey, William. "CPI Requirements of CE," Internal Bureau of Labor Statistics Paper, 2010. Available online at <http://www.bls.gov/cex/geminimaterials.htm>.

ror in the survey data, b) environmental changes related to spending behaviors, c) a need for greater flexibility in the mode of data collection, and d) the ability to modernize data collection. The new design of the CE must address these issues.

Reducing measurement error is the primary mission of the Gemini Project, yet as certain social behaviors and technological changes become more common, there is an increased likelihood of introducing measurement error to the survey. For example, purchases made online or recurring bill payments made by automatic debit may be less salient to respondents. Flexibility in CE data-collection strategies better positions the program to respond to such changes over the long term. Furthermore, while the spending behaviors that the CE seeks to measure have changed considerably over the past 30 years, the fundamental design of the survey has not. Although a number of improvements have recently been incorporated into the survey design, including the transition to a computer-assisted personal interviewing (CAPI) instrument in the Interview Survey, there has been no large-scale, comprehensive change to the survey design since 1980.

### **Gemini Project objectives**

The primary short-term objective of the Gemini Project is to develop a detailed redesign planning document. This document will describe the priorities, individual steps, timeframe, resource needs, and costs required to develop, pilot test, evaluate, and implement a redesigned CE. The redesign planning document will also guide the development and implementation of CE research studies throughout the Gemini Project lifecycle.

As previously noted, the long-term objectives of the redesign initiative are to introduce design changes in the CE that reduce measurement error, improve overall data quality, enhance the analytic value of CE data to users, and support a greater operational flexibility to respond to changes in the interviewing environment. These changes must occur within specified

budget levels. Therefore, all proposed changes will be subject to budgetary constraints, and implementation decisions will be considered in terms of priorities and trade-offs. To allow for an unpredictable budgetary environment, the redesign planning document will address both a complete redesign of the CE as well as more limited.

### **Gemini Project timeline**

A high-level project timeline includes information gathering in 2009 and 2010, to be followed by information synthesis and research planning in 2011. By the end of 2012, the project aims to have a proposed redesign planning document in place for the development and implementation of a redesigned CE. The development, testing, and implementation of a redesigned survey will occur in 2013 and beyond.

### **Redesign challenges**

Even at the outset of the survey redesign process, a number of challenges are evident. Defining survey requirements is challenging, particularly given the CE's diverse user community. Despite identifying the needs of varied users and reconciling competing interests, there are certain to be some users with unmet needs. Additionally, it is a challenge to gather, respond, and act on stakeholder concerns and suggestions while maintaining forward project progress. Finally, because of time constraints and uncertainty over the direction and nature of the forthcoming recommendations, the team is currently investigating several redesign topics, and this research will be used to support or reject some proposals. Ultimately, however, the redesign process is constrained by two overriding factors: the final survey design must produce the estimates required by the CPI, and, as mentioned above, long-term operational survey costs must keep to specified budget levels.

### **Project structure**

The Gemini Project is composed of five teams, all of which report to the Gemini Steering Team. (See exhibit 1.) Each

team addresses a specific objective that serves the overall Gemini mission. The Research Project Tracking System Team's primary objective was to create a research database for all completed, in-progress, and proposed CE research projects that could potentially inform the survey redesign. This team has finished its work, and the Research Project Tracking System is currently in use. The Data Quality Definition Team's objective is to produce a framework for assessing CE data quality, addressing both disparate user groups as well as key elements for an operational definition of data quality. The Data Quality Definition Team focused on six dimensions: relevance, accuracy, coherence, timeliness, accessibility, and interpretability.<sup>2</sup>

Three other teams were formed to plan, conduct, and summarize events that provide input and recommendations on issues related to the redesign. One of these teams—called the Data User Needs Forum Team—hosted an event in June 2010 where customers described their uses of the data, data requirements, priorities, and recommendations for changes. The team will summarize the findings from the forum and a subsequent user survey in a CE program office statement on CE data priorities.

The Conference Team held a Methods Workshop in December 2010. In addition, this team has organized a Survey Redesign Workshop where five Federal survey teams gave an overview of their redesign activities, and a Data Capture Technology Forum where Nielsen, National Opinion Research Center, RTI International, and Westat presented various technologies used to collect survey data. The Conference Team also organized a panel at the 2010 American Association for Public Opinion Research (AAPOR) on respondent record use.

Finally, the User Impact Team (which is not yet formed) is charged with assessing the impact of possible redesigns on data user products.

<sup>2</sup>Brackstone, Gordan. "Managing Data Quality in a Statistical Agency." *Survey Methodology*. vol. 25 no. 2, 1999, pp. 139-149.

## Exhibit 1. Gemini project structure

| Group                                 | Objective  | Outputs   |
|---------------------------------------|--|---|
| Gemini Steering Team                  | Oversee the Gemini Project   | Team charters, Gemini Project plan, redesign planning document  |
| <b>Teams</b>                          |  |   |
| Research Project Tracking System Team | Create a research database for all completed, in-progress, and proposed CE research projects that potentially inform the survey redesign | Research Project Tracking System  |
| Data Quality Definition Team          | Produce an operational definition of data quality for CE and a framework for assessing CE data quality by its disparate user groups      | Report summarizing a data quality definition and framework for CE   |
| Conference Team                       | Coordinate methodology events to solicit external input and recommendations on issues related to the redesign                            | <ul style="list-style-type: none"> <li>•Survey Redesign Panel</li> <li>•Data Capture Technology Forum</li> <li>•Respondent Records Use AAPOR Panel</li> <li>•CE Methods Workshop</li> </ul> |
| Data User Needs Team                  | Coordinate an event and survey to solicit input from data users on current use, requirements, and potential use of the data              | Report summarizing data user needs, priorities, and recommendations   |
| User Impact Team                      | Identify the impact of possible redesign alternatives on data user   | Report summarizing impact of proposed redesign alternatives on data users   |

After alternative designs have been proposed, this team will explore the impact of the redesign proposals on users and their uses of the data.

### Major activities

Several major products have been, or will be, created as part of the Gemini Project. The first is the Research Project Tracking System (RPTS) and accompanying report. The RPTS is a database that enables status tracking for all proposed, approved, in-progress, completed, and deferred CE methods research projects that have the potential to influence the CE redesign. The system stores project information in an organized structure, providing a mechanism for Gemini Project teams to better align current and proposed projects with specific redesign needs.

The second major product is the Data Quality Definition Team's report, which provides an operational definition of data quality for the CE program. The definition includes a statement of the procedures or ways in which the program should measure data quality, provides a framework for assessing the overall quality of CE survey data, and addresses the

data usability concerns of individual program stakeholders.

A third major product is a set of summary documents created after each Gemini Project event. These reports identify the presenters, key points, and implications of the event. For example, a document summarizing the Data Capture Technology Forum will serve as a starting point for future exploration into data collection technologies by identifying strengths and weaknesses of the technologies presented, and identifying possible avenues of additional research.

The User Needs Team will create a report summarizing the input provided by data users. This input, along with similar input from the CPI, will be used to create a list of CE priorities that identify the minimally required elements of CE data, as well as secondary data requirements. This is an important document for the Gemini Project because it will serve as a key reference for all subsequent events and discussions regarding potential design alternatives.

### Redesign topics

Topics for current discussion. Many different elements of survey methodology influence the plans of a survey re-

design. Listed below are topics deemed most important for current discussion in planning a redesigned CE. These topics are addressed first because they are fundamental to making later, more tailored decisions about issues such as mode, technology, recall aids, and overall respondent burden. Key issues associated with each topic, including benefits and risks of related design alternatives that need to be explored early in the redesign process, are also described.

*Global questions.* The Interview Survey averages 65 minutes.<sup>3</sup> The length is a result of both the breadth of expenditure categories included in the survey and depth of detailed information required for each category. The aim of the current Interview Survey is to collect information about every expenditure category from every household. One proposed approach to accomplish this objective (while also making the interview shorter) is to replace some of the detailed questions with global questions. Global questions ask about an expenditure category from an aggregate standpoint (for example, "How

<sup>3</sup>U.S. Department of Labor, Bureau of Labor Statistics. *BLS Handbook of Methods*. Washington: U.S. Government Printing Office: 2007

much did you spend on clothing?”), as compared with a detailed series of questions about individual expenditure types within the larger category (for example, “How much did you spend on shirts, pants, sweaters, vests, etc.?”). Replacing some of the detailed question sections with global questions would reduce the length of the survey while still providing high-level expenditure information about the given categories. However, the quality of data collected with global questions is debatable, and global questions have the potential to increase the cognitive burden of the respondent because of question difficulty. Both of these issues need to be further explored.

*Interview structure.* Currently the Interview Survey is a highly structured and standardized CAPI interview. The instrument is programmed to administer the questions in a set order, and the questions are designed to be read by the interviewer exactly as worded. The questions are organized into sections by topic (for example, housing, utilities, vehicle expenses, insurance, educational expenses, and so forth), a structure which forces respondents to report expenditures in a prescribed format, rather than allowing them to draw on information in a sequence that might be easier for them to recall. Nonstructured interviewing may hold promise for collecting higher quality data, but will require a very different approach to data collection. In addition, it is more difficult to administer and may require more skilled interviewers to ensure the collection of all required data.

*Proxy reporting.* Currently only one respondent reports expenditures in both the Interview Survey and Diary Survey, answering questions for the entire household. The accuracy of the data provided by the proxy depends on the extent to which the respondent has detailed knowledge about the expenditures of other household members. Given the complexity of many household situations, a proxy

respondent may often be unaware of, or unfamiliar with, purchases made by others in the household. One approach to eliminating proxy reporting is to collect information from each person, either through an individual diary or an interview. The impact on data quality and response rates, and the risk that the same expenditure will be reported by more than one household member, are all issues requiring further investigation.

*Recall period.* The Interview Survey currently asks respondents to report expenditures for the past 3 months. The length of this recall period, combined with the wide range of questions, may contribute to incomplete or less accurate reporting, and may also present a substantial cognitive burden for respondents. A shorter recall period may result in improved recall and therefore higher quality reporting; however, the potential increase in the number of interviews per household required to support annual estimates may negate the benefits of increased reporting accuracy or decreased respondent burden resulting from the change. An increase in the size of the sample could counter some of these drawbacks, but would incur a significant increase in survey costs and thus would not fit within the constraints presented for the redesign.

*Split questionnaire design.* Currently, all Interview Survey respondents are asked about the same set of expenditure categories during each interview. The resulting collected data are used to estimate average annual expenditures in the various categories at the household level. Covering the same set of expenditure categories with every respondent means that the interview is long and burdensome. It may be possible, however, to administer subsets of expenditure categories to separate subsets of respondents and still generate average annual household-level expenditure estimates. Calculation of these estimates may depend on the statistical capability to model the noncollected data (for imputation) or the willingness

of the survey stakeholders to permit other types of data manipulation, such as matching similar households.

### **Future topics**

In addition to the topics reviewed above, several other areas will be explored during the redesign process. Before a final redesigned CE can be developed, some topics should be addressed, including the use of administrative records or external data, mixed-mode designs, new technology, and recall aids.

*Administrative records/external data.* Some of the information that the CE collects is also compiled by other sources. Use of external data could reduce respondent burden and potentially improve data quality. Whether these data can be used to reduce the amount of data that the CE collects depends on the availability of procedures to link or match CE sample units to external data, the quality of those data, and respondent willingness to allow the linkage.

*Mixed-mode designs.* Offering more than one data collection mode has the potential to increase response rates and reduce nonresponse bias, although research evidence in this area is mixed. Because the Interview Survey was designed to be administered by personal visit, and a substantial percentage of cases are administered by telephone, redesigning the survey questions to be either mode-neutral or tailored more specifically to the mode of actual administration has the potential to improve data quality and reduce measurement error.

*New technology.* New data collection technologies, such as personal digital assistants (PDAs), smartphones, or various types of scanning technologies, may have the potential to improve data quality, reduce respondent burden, and increase completion or response rates. Such electronic data capture devices eliminate information transfer from paper diaries to database, instead keeping the data in electronic form through-

out the process. Additionally, the portability and potentially increased convenience of electronic data capture devices may lend an immediacy and ease of use to the recording process that may substantially reduce recall burden.

*Records and recall aids.* Increased reliance and improved guidance on records, receipts, and recall aids, including electronic records, also has the potential to improve data quality. If the CE is able to identify a process to extract data from existing respondent records, respondent burden would also be reduced.

### **Redesign-related research**

To help make informed decisions in the redesign process, a number of research projects have either been initiated or are in the planning stages. These redesign-related research activities are presented below.

*Current studies.* Several research projects directly address the effect of questionnaire length on data quality and response rates. It is often assumed that the longer the questionnaire, the lower the quality of the data, and the less likely respondents will be to complete the entire interview, either in the current wave or in subsequent waves. The literature provides mixed results on the relationship between questionnaire length and data quality or response rates, with no clear implications for the CE. Therefore, the CE is exploring this issue directly.

The Measurement Issues study has four goals, two of which are to assess the impact of a shorter interview with a split questionnaire design and the use of global questions. By dividing the Interview Survey into portions and administering each portion individually, the study will be able to draw some conclusions about the impact of questionnaire length, as well as the quality of data obtained via global questions instead of detailed questions. In addition to studying questionnaire length and the use of global questions, the Measurement Issues study will also investigate the effect of recall period

on data quality. The Interview Survey currently uses a 3-month reference period, which may be too long for respondents to reliably recall all types of expenditures; the test allows a comparison with data collected using a 1-month reference period. However, a shorter reference period may require more frequent data collection to allow for annual expenditure estimates, so understanding the impact of monthly data collection on response rates is the fourth goal of the study.

The Order Effects Test study also examines questionnaire length. This test will provide data on whether the order of the interview section (that is, earlier or later placement) has an effect on the quality of data. This test randomly assigns first wave cases to one of two conditions. The first condition will have respondents proceed through the standard Interview Survey interview. The second condition moves a section from the middle of the Interview Survey to an earlier stage of the interview. If moving a section earlier in the interview significantly improves data quality, the results would support efforts to shorten the interview.

Over the past several years, the CE has explored the use of split questionnaire methods as an alternative to the current data collection procedures. These methods involve dividing a questionnaire into subsets of questions and then administering each subset to a subsample of respondents. Currently, the feasibility of this type of split questionnaire design is being investigated as one of the test conditions in the Measurement Issues Study, described above. Serving as a counterpart to the Measurement Issues Study, ongoing analyses use historical CE data to run statistical simulations investigating the value of an allocation method to assign survey items to groups of households based on previous waves of data. Since split questionnaire methods do not ask all survey items of all households, additional research has focused on methods to “fill-in” this missing data. Specifically, the CE has explored the use of imputation-based procedures to estimate from all households instead

of only those households that were directly asked about a given expenditure category. Future research will look at refining the allocation and imputation procedures so that more efficient estimates can be obtained.

*Planned studies.* There are two other key issues that need to be addressed through research before major redesign decisions can be made. The first issue is the effect of proxy reporting. As noted above, the CE program relies on one respondent to provide expenditure information for the entire household, a procedure suspected of being a major source of measurement error from underreporting. The Individual Diary Study will investigate the feasibility and impact of collecting data from all household members. This study will provide an individual diary to each eligible household member. The collected data will be analyzed for improvements in data quantity and quality over current production data, as well as the impact of the procedural change on collection costs and response rates. An online diary component is planned as part of this test.

The second issue that the CE plans to address through research is measurement error. Comparison studies with external data, such as the Personal Consumption Expenditures, suggest that CE data may have significant measurement error, especially in certain categories. Currently, there is a limited understanding as to the sources of this error, and no estimates of the magnitude of the errors. The Records Study is planned to address both of these issues. In this study, participants will complete a partial CE interview. They will then be asked to locate all available records that are relevant to the expenditures covered in the interview. In a subsequent visit, interviewers will attempt to match the records with reported expenditures and to identify differences between the reported expenditures and those on the records. The comparison between reported expenditures and records-based information will provide estimates of the accuracy of the expenditure reports. A follow-up

in-depth discussion with participants will also provide insight into the causes of the errors (for example, forgetting, comprehension difficulty, proxy issues, and so forth).

## Summary

*Project accomplishments.* The Gemini Project is designed to gather information on alternative survey designs and methodologies, to investigate the impact and feasibility of various options and propose relevant research, and to develop a planning document for redesigning the CE over approximately 3 years. Much has been accomplished to date: project teams completed data quality and research tracking reports, and drafted issue papers on preliminary redesign topics (available online on the

Gemini Project website at <http://www.bls.gov/cex/geminimaterials.htm>). Additionally, in early 2010, several successful events were held to discuss issues related to the redesign, including a Survey Redesign Panel Discussion, a Data Capture Technology Forum, and a Data Users' Needs Forum.

*Future plans.* Looking ahead, the CE Steering Team is planning several major events and additional research studies. A methodological workshop focused on current research findings regarding interview structure, global questions, proxy reporting, recall period, and split questionnaire designs, and their implications for the CE redesign, was held in the fall of 2010. As an input to the methodological workshop, a

review of current survey methods used in international consumer expenditure surveys will be prepared. In 2011, BLS plans to work with a consensus expert panel formed by the Committee on National Statistics (CNSTAT) to coordinate two events and produce a report with redesign recommendations based on the event discussions and other outside independent proposals. The 2011 events will include a household survey data producer workshop designed to determine common challenges and best practices for large-scale household survey data producers, as well as a follow-up workshop to discuss options for the CE redesign, based on internal CE research, outside proposals, and input from the CNSTAT consensus expert panel. ■