

Gemini Project Vision Document

Gemini Project Steering Team

Bureau of Labor Statistics

April 16, 2010

Last Revised:

February 25, 2015

The purpose of this document is to provide an overview of the Gemini Project.

Gemini Project Mission

With the Gemini Project, the U.S. Bureau of Labor Statistics (BLS) aims to redesign the Consumer Expenditure (CE) survey to improve data quality through a verifiable reduction in measurement error—particularly error caused by underreporting. The effort to reduce measurement error will combat further declines in response rates by balancing any expected benefits of survey design changes against any potential negative effects on response rates. Any improvements introduced as part of the Gemini Project should not increase budgetary burden, but instead, should remain budget neutral.

CE Survey Objectives and Uses

By improving the quality of the CE estimates, the CE survey (comprised of the CE quarterly interview, or “CEQ,” and the CE diary, or “CED”) satisfy their primary purpose: maintaining the integrity of the expenditure weights used in the Consumer Price Index (CPI). Improved data quality and stable or increasing response rates enhance the usefulness of the CE data for both public and private users. In the Federal sphere, for example, the Bureau of Economic Analysis (BEA) uses the data in estimating Personal Consumption Expenditures (PCE), while the Bureau of the Census employs Consumer Expenditure Survey-based thresholds to produce experimental poverty statistics. Congressional committees also request special tabulations of the data on issues such as the potential impact of increases in the minimum wage. The academic community and research organizations use CE public-use microdata extensively in their analytical work, investigating topics such as consumption over a person’s lifecycle, trends in expenditure inequality, the impact of public policy proposals on spending by consumers, and the consumption behavior of consumer units given differences in socioeconomic and demographic characteristics. Private companies use both CE tabular data and CE microdata in marketing their products and services. They do this directly or through firms that purchase CE microdata to provide such analytical services. In addition, CE data are a recommended resource for prospective entrepreneurs developing their business plans.

Motivation for Redesign

The design of the CE survey must be updated on an as-needed basis to address the effect of changes in society, technology, new consumer products, and spending methods on survey estimates. Without these updates, the CE survey will not be able to continue producing high quality expenditure estimates for users.

Surveys today face well-known challenges that affect response rates, such as increasingly busy respondents, confidentiality and privacy concerns, competing surveys, controlled-access residences, and non-English-speaking households. The CE survey face unique challenges that directly influence the quality of the data collected. Presented in order of importance, the three most challenging issues for the CE survey are the evidence of measurement error in the survey data, environmental changes related to new technology and consumption behaviors, and the need for greater flexibility in the mode of data collection and ability to update data collection strategies.

1. Measurement Error. From 1984 until the present, the ratios of aggregate expenditure estimates from the CE compared with PCE data from the National Accounts show a declining trend for a large number of spending categories, serving as evidence of underreporting. Similar findings from internal methodological studies and the 2008 CE Program Review Report present further evidence of a growing concern about the quality of reported data. Underreporting in the CE may result from perceived (and real) respondent burden due to survey length and complexity, panel or questionnaire conditioning, increasing telephone administration of a survey originally designed for personal visit interviews, proxy reporting by a single household member, recall effects stemming from a 3-month reference period, and other sources of measurement error.
2. Environmental Changes. As new technologies evolve and are adopted, the CE survey must adapt to account for the environmental changes that accompany technological innovation. The two most important changes are the transformation of respondent behavior and the emergence of new data collection options. For example, online purchases and automatic payments are two popular methods of payment, and questionnaire design and data collection methods may have to be adapted to better account for them. The availability of new technology for data collection, such as a web diary, a portable digital assistant (PDA) instrument, or financial-software-generated balance sheets, offers new opportunities by which to collect data. Additionally, respondents may expect to be able to respond to the survey online, a service which BLS does not currently offer. Furthermore, a different survey design may better capture certain types of expenditures, such as

automatically-billed payments. Finally, the impact of shopping at large general merchandise stores, such as Costco and Walmart, on expenditure reporting is not known, but the topic-specific design of the survey may not correspond well to purchases of a large variety of goods at a single store. A different survey design may be better suited for capturing such expenditures.

3. Greater Flexibility. The CE needs greater flexibility for two main reasons: to enable CEQ interviewing in more than one mode and to allow for faster implementation of changes to the questionnaire or instrument. Regarding the first reason, a multimode design would allow data collection to be tailored to the needs of the respondent, and could help to address the environmental changes noted above. For example, respondents who have very little time, have difficulty keeping a CED diary, or do not want an interviewer in their home all have different ways of optimally reporting their data. Accordingly, they would be more likely to complete an interview using one mode over another. The second reason for greater flexibility is that, as currently designed, changing the computer-assisted personal interviewing (CAPI) instrument requires considerable lead time. Increased flexibility in the instrument would allow the program to be responsive to social and external changes without a major instrument restructuring.

Reducing measurement error is the primary mission of the Gemini Project, but environmental changes are an increasingly important mechanism through which measurement error may be introduced to the survey. And greater flexibility in the ability to update or revise data collection strategies better positions the survey to quickly respond to these environmental changes over the long term.

Furthermore, the consumer expenditures that the survey seeks to measure have changed considerably over the past 30 years, while the fundamental design of the survey has not. Although a number of smaller scale improvements have been incorporated into the survey design, including the transition to a CAPI instrument, these changes have not been made within a larger, more comprehensive framework of addressing identified survey weaknesses.

BLS initiated survey redesign plans several years in advance of the target implementation date in order to account for the length of time needed to develop, evaluate, pilot test, and implement a large-scale survey redesign. The overall timeline should not exceed 5 years for development and pilot testing, with a new survey in the field within 10 years.

Gemini Project Objectives

The short-term objective of the Gemini Project was to develop a detailed roadmap for the redesign of the CE survey, which will promote improved expenditure estimates through a reduction in measurement error. The completed redesign roadmap describes the priorities, individual steps, timeframe, resource needs, and costs required for the development, pilot testing, evaluation, and implementation of a redesigned CE survey. The roadmap also guides the development and implementation of CE research studies throughout the Gemini Project lifecycle.

The long-term objectives of the redesign initiative are to reduce measurement error, improve overall data quality, enhance the analytic value of CE data to users, and implement a new design that supports a greater operational flexibility to respond to changes in the interviewing environment.

Finally, the costs of the redesigned survey cannot exceed current budget levels. Therefore, all proposed changes are 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 planning document will address both a complete redesign of the CE survey as well as more limited modifications to the current design.

Approach

The Gemini Project is managed by a project leader and steering team, with guidance from an executive management group. The Gemini Project includes three stages of work. (The timeline of Stages 1 and 2 are outlined in Attachment 1.) Stage 1 focused on summarizing research to reduce measurement error in the process of redesigning the CE survey and consisted of five major steps:

1. Establishing a chartered steering team to oversee project activities
2. Creating a research project tracking system
3. Defining data quality for CE
4. Coordinating a series of information gathering events
5. Proposing redesign models

Multiple information-gathering meetings, conference sessions, forums, and workshops were conducted in Stage 1, including the following events:

- Survey Redesign Panel Discussion, co-sponsored by the Washington Chapter of the American Association for Public Opinion Research (DC-AAPOR) and the Washington Statistical Society (WSS) (January 2010)
- Data Capture Technology Forum (March 2010)

- AAPOR Panel on Respondent Record Use (May 2010)
- Data User Needs Forum (June 2010)
- CE Methods Workshop (December 2010)
- Consensus Expert Panel First Meeting (February 2011)
- Household Survey Producers Workshop and Consensus Expert Panel Second Meeting (June 2011)
- CE Redesign Options Workshop and Consensus Expert Panel Third Meeting (October 2011)
- Consensus Expert Panel Fourth Meeting (January 2012)
- Survey Methods Symposium (July 2012)
- CNSTAT Report Workshop (scheduled for October 2012)

Subsequent events have focused on the proposed redesign, as well as issues identified during past events and other survey methods topics of interest.

The following significant reports and contracts were also completed or awarded during Stage 1:

- Data Quality Definition Report (October 2009)
- CPI Requirements Paper (June 2010)
- Conference Team Report (August 2010)
- CNSTAT Consensus Expert Panel contract awarded (September 2010)
- CE Data Requirements Report (May 2011)
- Independent Redesign Proposals II contract awarded (June 2012)¹
- Gemini Technology Team Report (June 2012)²
- Gemini Sampling Team I Report (September 2012)³
- Small-scale Redesign Project Report (September 2012)⁴

¹ A contract with a vendor was awarded for the development and proposal of a CE Redesign given updated CE design decisions, cost estimates, and preliminary reconciled expenditure categories. Their first proof-of-concept presentation will be in November 2012.

² This team comprised of BLS and Census members reviewed the research literature to-date and recommended the most promising technologies for the CE Redesign Proposal.

³ This team estimated sample sizes required for the CED and CEQ based on a preliminary list of reconciled expenditure categories.

⁴ The Gemini Steering Team met during Winter and Spring of 2012 to discuss small-scale changes that could be made to CE. The goals of these changes were to reduce respondent burden by streamlining the interview, reduce costs, and to improve data quality (e.g., measurement error) to the extent possible without increasing burden or cost. A memo of suggested changes for 2015 was delivered and a memo for suggested changes for 2017 will be generated.

- OPLC Requirements Team Report (expected October 2012)⁵
- Independent Redesign Proposals II proof-of-concept presentation (November 2012)

The team that developed the CE Resign Proposal, the Gemini Design Team, was a joint BLS-Census team which kicked-off in July 2012 and submitted their proposal in July 2013, at which time it was also approved by BLS management.

Stage 2 integrated research findings from Stage 1 into a coherent plan consisting of two major steps: 1) Assessing user impacts and 2) Finalizing the redesign roadmap for piloting, evaluating, and transitioning to the new design. The redesign roadmap was completed in September 2013, and a Data User Impact report was released in February 2014.

Pilot testing the proposed redesign, conducting additional research on alternative designs, and implementing the approved redesign can be considered a 3rd stage of the redesign process. Stage 3 includes work to develop, test, and implement the survey redesign using the redesign roadmap approved by senior management. More detail on current and upcoming projects that are included in stage 3 can be found in the Gemini Research Agenda (http://www.bls.gov/cex/gemini_web_report.pdf), which will be regularly updated.

⁵ This team reconciled CE-CPI data requirements into one set of reconciled expenditure categories.

Redesign Topics

Many different elements of survey methodology affect a survey redesign. Table 1 includes a list of those topics that were most critical for preliminary discussions for a redesigned CE, a description of each, and examples of the forums where the topic were discussed. Some topics, such as administrative records or other external data, have been researched and determined to be sufficiently infeasible that CE will not continue to investigate them. In the case of administrative records, this decision was made due to concerns about complexity from multiple data collection streams, resource constraints, quality of administrative data, respondent compliance, source compliance (e.g., store or third party aggregator), representativeness, and other risks due to loss of control over collection. Note that Table 1 is not meant to be a complete list of topics, nor is it a complete list of venues for discussion.

More information about in-progress and planned BLS research projects to support the survey redesign can be found on the Gemini Project website at <http://www.bls.gov/cex/geminimaterials.htm>. Since these projects span a number of the survey methods topics, it made sense to list them separately rather than fitting them into this table.

Table 1. Topics for Investigation and Further Research (listed alphabetically)

Topic	Description	Forum for discussion
Administrative records/external data	Some of the information that CE collects is collected in other sources. Using external data could reduce respondent burden and potentially improve data quality (depending on the quality of the external data).	<ul style="list-style-type: none">- Household Survey Producers Workshop (June 2011)- CE Redesign Options Workshop (October 2011)
Data user needs	It is important to get input from data users to understand their uses of the data and to get suggestions for changes to the surveys.	<ul style="list-style-type: none">- On an ongoing basis, keep users informed of the status of the project through meetings and presentations, including the expert panel- Data User Needs Forum (June 2010)
Mixed-mode designs	Offering more than one data collection mode (e.g., telephone or web) may increase response rates and reduce bias. BLS must determine how to design and implement a successful mixed-mode survey.	<ul style="list-style-type: none">- Household Survey Producers Workshop (June 2011)- CE Redesign Options Workshop (October 2011)

Topic	Description	Forum for discussion
New technology	BLS will examine technologies for data collection, such as PDAs, smart phones, or various types of scanning technologies, and determine their possible use and cost for CE.	<ul style="list-style-type: none"> - Data Capture Technology Forum (March 2010) - Westat's "Data Capture Technology" presentation at FedCASIC (March 2011) - Household Survey Producers Workshop (June 2011) - CE Redesign Options Workshop (October 2011) - Survey Methods Symposium (July 2012)
Proxy reporting	Collecting data from all consumer unit members may result in better reporting. BLS will determine how best to collect data from all consumer unit members without increasing unit nonresponse.	<ul style="list-style-type: none"> - Household Survey Producers Workshop (June 2011) - CE Redesign Options Workshop (October 2011)
Questionnaire design	The use of alternative designs for the CE survey may allow for more efficient data collection or the collection of higher quality data. BLS will determine how CE can incorporate different types of interview structures or technology-driven approaches into the data collection process.	<ul style="list-style-type: none"> - Survey Redesign Panel Discussion, sponsored by DC-AAPOR and WSS (January 2010) - CE Methods Workshop (December 2010)
Recall period	A shorter recall period may result in better recall and therefore higher quality reporting. On the other hand, contacting respondents more often (such as monthly) might lead respondents to stop reporting or to report fewer expenditures.	<ul style="list-style-type: none"> - CE Methods Workshop (December 2010)
Records and recall aids	Increased reliance or improved guidance on records, receipts, and recall aids, including electronic records, may result in better reporting. BLS will determine how these resources can be integrated effectively into the CE.	<ul style="list-style-type: none"> - AAPOR Panel on Respondent Record Use (May 2010)

Topic	Description	Forum for discussion
Required data elements and features in redesigned survey	A formal process to specify the data elements, both expenditure and nonexpenditure, and ancillary features, including timeliness and periodicity, necessary to satisfy the needs of the CPI, CE and DPINR programs. This will become a requirements document to guide design teams on the data the redesigned CE survey should obtain.	- OPLC Requirements for CE Team (October 2012)
Respondent burden	An interview that is shorter or less burdensome cognitively could result in better reporting, higher retention rates, and reduced bias in later interviews. Possible approaches might include increased use of screener questions or asking subsets of questions supplemented by global questions in place of the whole interview.	- CE Methods Workshop (December 2010)

Criteria for Success

The success of the Gemini Project will be judged based on the accomplishment of the following criteria, listed in descending order of priority:

1. There is a measurable reduction in measurement error.
2. There is an overall increase in data quality, including a balanced evaluation of the mean squared error of the cost weights, and reduction of components of bias and variance attributable to specific sources of sampling or nonsampling error, such as unit nonresponse, wave nonresponse, or item nonresponse. Data quality criteria may also include other factors such as timeliness. Practical decisions on methodology and field operations will involve trade-offs among cost and error factors for different components of bias and variance. Realistically, it is not expected that there will be uniform improvement in all components of error: some will decrease, and some will increase.
3. The analytic value of CE data for researchers and policymakers is maintained or improved, both for microeconomic and macroeconomic data users.

To the extent possible, it is also important that the new design supports a greater operational flexibility to respond to changes in the interviewing environment, and that results from the success criteria are reasonably robust against future changes in societal and technological factors and variability in budgetary levels.

Stakeholders

The stakeholders for the Gemini Project include all groups involved in or concerned about the impact of survey design changes on the collection, processing, editing (including imputation and allocation), weighting, estimation, evaluation, quality, and use of CE data. These groups include:

- BLS Associate Commissioner, Office of Prices and Living Conditions (OPLC)
- BLS Division of Consumer Expenditure Information Systems (CEIS)
- BLS Division of Consumer Expenditure Survey (DCE)
- BLS Division of Consumer Prices and Price Indexes (CPI)
- BLS Division of Price and Index Number Research (DPINR)
- BLS Division of Price Statistical Methods, Branch of Consumer Expenditures (SMD)
- Bureau of the Census (BoC)
- Other Federal Agencies including the Bureau of Economic Analysis (BEA), the Internal Revenue Service (IRS), and the Department of Defense (DoD)
- Other individual users of CE microdata and published estimates in the academic, Government, nonprofit, and commercial sectors.

Attachment 1. Gemini Project Timeline

The current high level timeline is maintained on the Gemini website, accessible at <http://www.bls.gov/cex/geminitimeline.pdf>.

Below is a list of past completed tasks for Stages 1 and 2 of the Gemini project.

2009

February	Milestone	- Gemini Project kick-off
March	Milestone	- Research Project Tracking System Team, Redesign Conference Team and Data Quality Team chartered
July	Milestone	- Expert Panel chartered
August	Milestone	- Expert Panel kickoff meeting
October	Deliverable	- Data Quality Definition report completed
December	Deliverable	- CE Methods Workshop issue papers completed

2010

January	Event	- Survey Redesign Panel Discussion featuring representatives from five major Federal surveys
March	Event	- Data Capture Technology Forum
May	Event	- AAPOR (American Association for Public Opinion Research) Panel on Respondent Record Use
June	Event	- Data User Needs Forum
August	Deliverable	- Data User Needs Team report completed
August	Deliverable	- Conference Team report completed
November	Deliverable	- Statement on CE Data Priorities released
December	Event	- CE Methods Workshop
December	Deliverable	- Research Project Tracking System report completed

2011

June	Event	- Household Survey Producers Workshop
October	Event	- CE Redesign Options Workshop

2012

June	Deliverable	- Gemini Technology Team Report
July	Event	- CE Survey Methods Symposium
August	Deliverable	- CNSTAT Consensus Expert Panel pre-publication report
September	Deliverable	- Gemini Sampling Team I Final Report
October	Deliverable	- Public release of the CNSTAT Consensus Expert Panel report
October	Deliverable	- OPLC Requirements Team Report
October	Event	- CNSTAT Consensus Expert Panel Report Workshop
Fall and Winter	Milestone	- Develop models for redesign
December	Milestone	- Presentation at Federal Economic Statistics Advisory Committee Meeting

2013

July	Deliverable	- Gemini Design Team proposal completed
July	Milestone	- Presentations of the Redesign Proposal to CE Management for review and approval
July	Event	- CE Survey Methods Symposium and public release of the design plan
July -December	Deliverable	- Detailed redesign roadmap & presentations for proposed redesign

2014

April	Deliverable	- User Impact Team report published
-------	-------------	-------------------------------------