

Templates for Evaluating Survey Programs

Brenda G. Cox and Nancy J. Kirkendall¹

The Performance Assessment Rating Tool (PART) was instituted to encourage rigorous performance assessment to boost the quality of Federal government programs. For survey programs, an important component of performance assessment is the evaluation of each of the program's survey components and then the evaluation of the effectiveness of the overall program. As a part of its performance assessment, the Energy Information Administration (EIA) is developing two templates, one for the evaluation of an individual survey and one for the evaluation of a family of surveys that form a survey program. These templates will serve as a guide for an external group to use to conduct survey program evaluations. The templates are being tested by using the draft templates to evaluate the Petroleum Marketing Program and its component surveys. When this process is completed, the templates will be updated to create the penultimate templates for survey and program evaluations. This paper describes the two templates that have been prepared and the results of testing.

Keywords: PART, Performance Assessment, Survey Evaluation

1. Background

The Office of Management and Budget (OMB) is assessing the effectiveness of Federal agencies using the Performance Assessment Rating Tool (PART).² PART was instituted to boost the quality of Federal programs by encouraging rigorous performance assessment for government programs. Performance measures and goals are heavily emphasized, as required by the Government Performance and Results Act of 1993. The overall ratings are effective, moderately effective, adequate, results not demonstrated, and ineffective. OMB is assessing about one-fifth of all Federal agencies each year. Performance Institute studied the 2004 budget ratings and reported in *GovExec* that government programs with a rating of "effective" had a six percent budget increase while programs with a rating of "results not demonstrated" only had a one percent increase.

The PART consists of 25 questions grouped into four areas:

- Program purpose and design
- Strategic planning and performance measures
- Program management
- Program results and accountability

Two questions explicitly ask for evaluations. Question 2.6 asks:

Are independent evaluations of sufficient scope and quality conducted on a regular basis or as needed to support program improvements and evaluate effectiveness and relevance to the problem, interest, or need?

Question 4.5 asks:

Do independent evaluations of sufficient scope and quality indicate that the program is effective and achieving results?

The guidance specifies that the evaluations must cover the entire program—all activities of the Energy Information Administration, for instance. They must also contain recommendations for improvement and address the effectiveness of the agency. In particular, compliance requires recurring, independent evaluations that are program wide and that address achievement of performance targets and effectiveness.

Program strengths and weaknesses are identified in these performance assessments resulting in recommendations for improvement and recognition of desirable design elements already in place. For the Energy Information Administration, an important component of its activities is the production and dissemination of data through statistical reports and economic models. These products depend upon survey data collected via families of related surveys. For example, the Petroleum Marketing Program is composed of 11 surveys that monitor—at national, regional, and state levels—the quantities of crude oil and its finished products present at various market stages and the associated acquisition costs or prices being charged. To evaluate the effectiveness of

¹ Dr. Cox was supported in conducting this research under a subcontract Battelle has with Z, INC. for Task Order 04-EI-40503.000 from the Energy Information Administration (EIA). The views expressed are the authors' and do not necessarily reflect the official position of the Energy Information Administration.

² See www.whitehouse.gov/omb/part for further details.

programs that involve a family of related surveys, the individual surveys that comprise the survey program need to be evaluated as well as the effectiveness of their integration to create the survey program. This paper presents results of an investigation for the Energy Information Administration that developed templates for performing an external evaluation of a family of related surveys and its component surveys. The intention was to produce a Program Evaluation Template and a Survey Evaluation Template which could be used as guides for performing expert assessments of survey programs and their component surveys. The templates were proof of concept tested on the Petroleum Marketing Survey Program—a relatively stable, relatively well-documented family of surveys. The paper also highlights insights gained in the proof of concept testing process.

2. Evaluation Template for an Individual Survey

Program Evaluation for a family of surveys must begin with a summary evaluation of the individual surveys that comprise the survey program. Understanding the surveys that comprise a program is essential to assessing the effectiveness of their integration to create the survey program. The template for an individual survey will facilitate evaluation of these key components:

- survey objectives, including population subgroups for which separate estimation is desired;
- target population definition, exclusions and inclusions, and temporal dimension;
- sampling frame, including coverage of the target population and other quality issues;
- sample design characteristics and sample size allocation;
- data collection steps including outcome codes and response/eligibility results;
- forms design for screener and questionnaire;
- edit and imputation procedures for missing data items and/or data records;
- weighting, including nonresponse adjustments and checking; and
- data analysis methods, end products, and documentation.

In using the template to evaluate a survey, the reviewer summarizes the program's attributes with respect to each area, identifying strengths and weaknesses, and then providing recommendations for improvement and recognition of desirable design elements already in place.

3. Program Evaluation for a Family of Surveys

The Program Evaluation for a family of surveys begins with a brief description of the survey program and the individual surveys contained in the family of surveys composing the survey program. Next, the following topics are evaluated for the entire program:

- Program objectives
- Target Populations
- Sampling frames
- Program Design
- Data Collection and Processing
- Data Analysis, Products, and Documentation
- Summary of Findings

In the course of completing the Program Evaluation, the reviewer will examine:

- which respondents report to which surveys, and the data requests made—to identify unnecessary burdens placed on survey respondents and whether the survey program needs restructuring;
- how the data within and across surveys are used in analysis and presented to the public and the relationship to the survey designs—to determine if one or more of the surveys should be modified;
- the relationship between question wording and consistency from EIA's and the respondent's points of view—to decide if there is a better design for forms and instructions;
- how well confidentiality protection is applied to assure confidentiality and encourage response; and
- response issues and timeliness.

The program evaluation will identify strengths and weaknesses of the program and will result in recommendations for improvement and recognition of desirable design elements already in place. In addition to its use for a particular survey program, the methodology being developed could also be extended to do evaluations across survey programs that have related topic areas and populations such as electric power with natural gas, and petroleum supply with petroleum marketing.

4. Template Testing

The program and survey evaluation templates were concept tested by using them to evaluate the Petroleum Marketing Program. The evaluation was based on publicly available documentation, including the most recent OMB clearance package, published

data reports, and EIA web site postings. The completed evaluations were then shared with program and survey managers to get their comments or corrections. As appropriate, the input from program staff has been incorporated into the final survey and program evaluations.

At this point in time, the survey evaluation template has been used to review 6 of the 11 petroleum marketing surveys. The review resulted in a 7 to 11 page summary for each survey with specific recommendations for improvement. Typical findings for the survey evaluations were that the survey objectives need to be more completely specified, the target population definitions should be made more specific, frame coverage and updating procedures were not adequately specified, editing and imputation procedures needed to be more fully documented, and the mailout packages are well designed but could benefit from tweaking.

Several important results came out of this process. The most important finding was that the survey evaluation process worked. The underlying concept for the survey review was that the reviewer would first summarize the methodology for each topic area in the template, then critique the survey approach and provide recommendations for change or endorsements of the current approach. When EIA staff reviewed the draft survey evaluations, however, we discovered that the recommendations need to be explicitly labelled and presented on a section by section basis. Thus, recommendations associated with the sampling frame would be presented in the sampling frame section. Another finding was that the same design attributes and flaws often occurred across surveys. Each survey evaluation was independently prepared so this repetition made for boring reading for EIA staff reviewing the whole package. Finally, the survey evaluations were impeded by the absence of survey methods reports that overviewed all aspects of survey design and execution. Instead, a variety of sources had to be used to evaluate each individual survey, and some aspects of the survey could not be evaluated.

The program evaluation template has been used for the overall evaluation of the Petroleum Marketing Program. One finding from the concept testing was that the survey evaluations were critical to the evaluation of the overall program. Part of the reason for their importance was the discovery that there was essentially no documentation describing the overall program. The available documentation focuses on the individual surveys and does not address how these surveys come together to create a program of research. The individual surveys have evolved over time and many were inherited from other agencies. As a consequence, variations occur in target populations

and in reporting procedures/definitions that appear to be due to happenstance as well as to deliberate planning.

Recommendations resulting from this evaluation tended to be related to the need an overall design plan. These recommendations revolved around three themes: (1) the need for a conceptual plan for the overall program describing its objectives and specifying how the individual survey components relate to these objectives; (2) the need for the individual surveys to tighten up some definitions/procedures to eliminate unnecessary variation across surveys; and (3) the need for methods documentation at the program level and at the survey level.

The template form can be used as a guide for what should be included in the methods documentation. An additional possible use of the completed templates is as a starting point for the documentation for a particular survey or survey program.

5. Remaining Activities

Currently, we plan to evaluate the remaining five surveys in the Petroleum Marketing Program, which will provide an opportunity to perfect the survey evaluation process. These five survey evaluations will be used to revise the draft program evaluation and the package will again be given to survey staff for their review. Once comments are received, the evaluations will be corrected and a Final Evaluation Report will be prepared.

The final activity in this investigation will be to use what was learned in evaluating the Petroleum Marketing Program to create penultimate templates for the survey evaluation and the program evaluation. These templates and their use will be documented in a methodology report with details on their use.

Appendix A: Survey Evaluation Template

Survey Name:

Name of survey and (if applicable) the family of surveys to which it belongs.

Brief Description of Survey:

Include subject matter, sponsors, and periodicity (one-time, repeated cross-sectional, or longitudinal).

Survey Objectives:

Describe the primary purpose for the survey. Compile a detailed list of the uses for which the data are to be input and the research questions that the data must answer. Indicate whether the survey is primarily intended for descriptive purposes versus policy research. List reporting domains for which separate estimation is needed. Indicate when and to what extent comparisons will be made to other surveys or to previous cycles of the same survey.

Target Population:

Delineate the entire set of population units for which the survey data are to be used to make inferences. If business entities are involved, indicate the level of the business forming the units of the target population. If inferences are desired for multiple population units, then document each unit. Be specific so that inclusions and exclusions from the target population can be completely understood. Specify the temporal and geographic aspects of the target population units. Specify the treatment of units whose characteristics change over time. Describe the treatment of units that enter the population (births), exit it (deaths), or assume a new identity (movers, ownership changes, mergers).

Sampling Frame:

Describe the list or mechanism used to enumerate population units for sample selection and data collection purposes. Specify differences between the units comprising the target population and the units comprising the frame. Delineate the relationship between frame units and target population units (one to one, many to one, one to many). Discuss whether duplicate listings are present and how treated in frame building. Discuss the extent to which target population members may be missing from the frame and what is known about the characteristics of the uncovered units. If appropriate, discuss size measures available/constructed for the frame and their accuracy.

Sample Design:

Note whether the survey is based upon a sample or a census of population members. For samples, provide an overview of the design, including whether selection procedures involve explicit or implicit stratification, the sample selection method(s), any multiple stages or phases of selection, definition of strata, use of size

measures, and unique or unusual approaches. Present the rationale for the sample design chosen.

Sample Allocation:

Summarize the sample size allocation to strata and the methodology involved. Describe the precision constraints/guidelines used to determine sample size in terms of completed interviews. List assumptions made about response and eligibility rates that were used in converting the sample allocations into the number of cases to be selected from each strata.

Data Collection:

Summarize steps in data collection. Review outcome codes in terms of capture of available information about nonrespondents based upon where nonresponse occurs in the interview process. Review response and eligibility results for the survey—tabular display best with outcome codes and formulas given. Screener and questionnaire review in terms of target population definition, capture of data needed for weighting, structure of form, and instructions. Review of mailout package materials if pertinent.

Editing and Imputation:

Review data cleaning and editing methods. Document relative magnitude of data records changed in processing or set to missing and whether too much/not enough editing has been done. Review imputation procedures in terms of method(s) used, their appropriateness, and the extent of missing data replaced. Document QA procedures implemented to ensure data quality.

Weighting:

Summarize steps in weighting and treatment of nonresponse/undercoverage. QA procedures implemented during weighting. Agreement between weight totals and external data sources.

Data Analysis:

Review reports to determine uses made of the data and domains of particular importance based upon frequency of use. Review other uses that are being made of the data, such as using the data as input for model building. Relate findings back to sample design and data capture.

Summary of Findings:

Summarize the findings of the evaluation with respect to the ways in which the survey is incorporating appropriate survey methods and/or creative or innovative procedures. Summarize any areas where the survey departs from what would be considered best practices and provide recommendations for improvement

Appendix B: Final Program Evaluation Template

Program Name:

Name of program.

Brief Description of Program:

Include program mission, subject matter, sponsors, reporting cycle. List individual surveys within the program and their relation to the overall program.

Program Objectives:

Describe the primary purpose for the program. Describe customers for program end products. Record legislation under which program activities are conducted and intended goals. List uses for program data and the role each component survey plays. Indicate whether the program is primarily used to satisfy descriptive needs versus modeling or policy research. Indicate when and to what extent data are being/should be integrated across surveys in data analysis.

Target Population:

Delineate the entire set of population units for which the program collects data. If business entities are involved, indicate the relationship between the entities across survey components, including population overlaps. Be specific so that inclusions and exclusions from the program's overall target population can be completely understood. Specify the temporal and geographic aspects of the target population units. Specify the treatment of units whose characteristics change. Describe variations in population definitions across surveys.

Sampling Frame:

Describe the lists or mechanisms used to enumerate population units for individual surveys and the extent of data sharing/coordination across surveys. Specify differences between the units comprising the program's target population and the units comprising the survey frames. Delineate the relationship between frame units across surveys. Explore how coordinated frame development is across surveys and to what extent there is data sharing/coordination across surveys of deaths, births mergers, and other changes of entities.

Program Design:

Review the conceptual design for the overall program, the rationale for the survey components comprising the program, how the surveys fit within the program's data collection plan, and their adequacy for their intended role. Delineate how well the component surveys come

together to create a unified, cohesive program design. Examine the reporting cycles across survey components and their relation to data integration. Explore potential population overlaps across survey components and any unnecessary/unusual burden placed on some population subgroups. Describe desirable design aspects of the program and its component surveys and aspects of the program that need restructuring.

Data Collection and Processing:

Summarize commonalities underlying data collection procedures/definitions across surveys as well as instances of intended/unintended variations across surveys. Review response rate patterns across surveys and their impact on programmatic reporting reliability. Consider the relationship between question wording and consistency across surveys from the respondent's point of view. Evaluate data collection impact for large businesses and how process is/should be tailored to meet their special requirements, whether within the same program and across programs. Determine if there is appropriate commonality of definitions and procedures across surveys and whether any data items are being collected repetitively across surveys. Review confidentiality assurance procedures for data capture and reporting.

Data Analysis, Products, and Documentation:

Review reports to determine how each survey component contributes to the whole. Review other uses being made of the data such as for model input. Relate findings back to relative importance of each survey component and program goals. Evaluate quality and understandability of data reports. Comparative analysis of data gathered against data from other sources such as other EIA survey programs, other Federal agencies, and private information from publications and associations. Explore resources devoted to quality analysis and control such as when respondents are contacted to explain differences between data sources.

Summary of Findings:

Summarize the findings of the evaluation with respect to the ways in which the program is incorporating appropriate survey methods and/or creative or innovative procedures. Summarize any areas where the program departs from what would be considered the best integrated design. Summarize data confidentiality procedures for reporting.