

Background Material for PART Evaluations

Thursday afternoon, Oct. 28 at 2:40 p.m.

OMB is assessing the effectiveness of federal agencies using the Program Assessment Rating Tool (PART). Performance measures and goals are heavily emphasized, and this is built off of the Government Performance and Results Act (GPRA) of 1993. The overall ratings are: effective, moderately effective, adequate, results not demonstrated, or ineffective. During the first year, the largest category was “results not demonstrated” with 50%, and only 5% were rated “ineffective.” The relation to an agency’s budget is not guaranteed, but a study of the 2004 budget ratings by the Arlington, Va.-based think tank Performance Institute and reported in GovExec showed a 6% increase for programs rated effective and a 1% increase for those with “results not demonstrated.” OMB is assessing about one-fifth of the agencies each year, and this, the third year, included us.

The PART consists of 25 questions grouped into four areas:

- Program purpose and design – we did fine here with 100%,
- Strategic planning and performance measures, – we were mortally wounded here,
- Program Management – fine with 100%,
- Program results and accountability – more problems.

EIA is in the process of switching our measures, and our examiner agreed that our new measures are more appropriate than our old. But since we did not have baselines or targets for the new measures, we received a “no” for several questions of having performance measures and targets, which mandated a “results not demonstrated” for the assessment as a whole. Our lack of evaluations is a problem we need to address, but it was not the cause of our overall rating.

Evaluations are asked for in two places. Section 2 deals with strategic planning, and 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?

The guidance covers several pages, and includes that evaluations must cover the entire program (i.e. all of EIA), contain recommendations for improvement, and address the effectiveness of the agency. Guidance on effectiveness drives at what is different in the world with your agency in existence. The results must also be used in planning.

Our answer mentioned the ASA Energy Committee, our independent Expert Reviews, and that our models are in the public domain. OMB stated that we have no recurring independent evaluation of sufficient scope (program-wide and addressing achievement of performance targets), effectiveness, or independence. Since no partial credit is allowed on this question, we received a NO rating.

Section 4 deals with program results and accountability, and question 4.5 asks:

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

The guidance says that the results of the evaluations mentioned in 2.6 can be used here, provided that they are of sufficient scope and quality.

Our evidence included the ACSI customer survey, and a study of the impact of our data on markets. OMB said “while EIA appears to be effective and achieving results, based in large part on universal reliance on EIA’s data and analysis, no independent evaluations of the entire program have been conducted.” We received partial credit with a rating of SMALL EXTENT.

The PART evaluations appear to be an evolving process, and our examiner admitted that the criteria may be more stringent this year than it was in the past. But the gold standard so far for statistical agencies appears to be BLS which cites the Boskin Commission report and follow-up studies. M.J. Boskin and four other Professors were commissioned by the Senate Finance Committee in 1995 to investigate whether the CPI was over-stating inflation. Some argue that they were chosen for being known critics of BLS. They reviewed already existing studies (many conducted within BLS) and issued a report in December 1996, which included some specific details. BLS has stated for years that the CPI is not a true cost of living index, but did implement many of the recommendations. Follow-up studies were conducted by ASA and by the National Research Council.

Although our rating is not yet posted, the OMB website has general information, guidance, and ratings conducted in prior years at www.whitehouse.gov/omb/part.

Attachment: Program Evaluation

Here are highlights from an 18 page pdf on program evaluation, with complete file at www.whitehouse.gov/omb/part/2004_program_eval.pdf

Question 2.6 asks whether there 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.” The purpose of question 2.6 is to ensure that the program (or agency) conducts non-biased evaluations on a regular or as-needed basis to fill gaps in performance information. These evaluations should be of sufficient scope and quality to improve planning with respect to the effectiveness of the program.

Question 4.5 asks if “independent evaluations of sufficient scope and quality indicate that the program is effective and achieving results.” The purpose of question 4.5 is to determine whether the program is effective based on independent and comprehensive evaluations. This question may be particularly important for programs that have substantial difficulty formulating quantitative performance measures. This question suggests that the quality of program evaluations presented in question 2.6 be strongly considered in answering this question.

.....

What sorts of tests provide strong evidence of a program’s effectiveness?

One of the central challenges in developing strong evidence of a program’s effectiveness is valid measurement of the difference between (i) the outcomes when the program is in place, and (ii) what the outcomes would have been in the absence of the program.

In cases where it is not possible to use an RCT (Randomized Control Trial) to evaluate the effectiveness of a program intervention, other approaches may be needed to evaluate: *What difference does the program make?* To approach an assessment of impact, the analysis must make every effort to compare the effect of the program with a baseline of what would have occurred in the absence of the program—an extremely difficult test. Finally, if it is not possible to evaluate the impact of a program, other evaluation approaches may shed light on *how* or *why* a program is effective (or ineffective), or may provide other information that is needed for the management of the program.

Here are relevant sections from FAQ’s at www.whitehouse.gov/omb/part/2004_faq.html

Program Evaluation

22. What should we consider when determining whether an office or official is sufficiently “independent” to conduct a program evaluation (PART Question 2.6)?

In general, for an evaluation to be independent it should be conducted by a group which is separate from those who administer the program.

What if the agency or program has its own evaluation office?

If the agency has an evaluation office that is an administratively separate unit from the program, the office will likely be sufficiently “independent” to conduct an

evaluation of the program. An evaluation conducted by the office administering the program would generally not provide the kind of independence needed to ensure that there is no conflict of interest.

What about programs administered by contractors?

If the contractor is administering the program on behalf of the program office, neither the program office overseeing the contractor nor the contractor itself would be considered "independent" for purposes of conducting an evaluation of the program's effectiveness.

What about an Inspector General? ... the inspector general probably has sufficient "independence" to conduct an evaluation of the program.

23. How do I demonstrate impact through a program evaluation?

The "randomized controlled trial" represents the best means to assess the impact of programs. However, this approach may not be feasible for all programs. ...

24. How do we determine what is the appropriate method for evaluating the program (PART Question 2.6)? If we cannot demonstrate the program's "impact", can we still receive credit for PART Question 4.5?

... The key question to answer is whether there is a form of evaluation that can provide strong evidence of a program's impact. If not, then consider the strongest evaluation possible that sheds light on the program's impact--even if only answering how or why a program is effective (or ineffective). The bottom line, though, is that the evaluation must be appropriate to the type of program. ...

If it is not possible to demonstrate impact, you can still receive credit for Question 4.5 if there are appropriate measures in place to show effectiveness. Credit can be given for evaluations that help determine *how* or *why* a program is effective (or ineffective). ...