Public Issues Forum

The Program Assessment Rating Tool (PART): What PARTs Help, and What PARTs Don't

First Annual AEA Public Issues Forum
Thursday, November 2, 2006

AMERICAN
EVALUATION
ASSOCIATION

This publication contains edited proceedings from the first Public Issues Forum sponsored by the American Evaluation Association at *Evaluation 2006: The Consequences of Evaluation*, held November 2, 2006 in Portland, Oregon.

Printed in the United States of America.

Table of Contents

Introduction

Welcome to the proceedings of the first American Evaluation Association Public Issues Forum. The AEA Public Affairs Committee, led by current committee chair and AEA President-Elect Bill Trochim, developed the idea of the Public Issues Forum as a new and potentially useful mechanism for engaging AEA in addressing timely and important issues for our field.

A public issues forum is an AEA panel designed to encourage constructive dialogue and to engender awareness among leaders, policy-makers and the general public on issues of pressing importance regarding the role, image or implementation of evaluation in public contexts. The forums are expected to encompass a diversity of views on the selected issue, to address the issue in a civil manner, and to be an educational endeavor, striving to present nuanced and thoughtful multiple perspectives in order to achieve deeper understanding. The topic that the forum addresses is selected for its relevance and importance to the field of evaluation, for its immediacy, and because it is at the interface of evaluation and the public.

The AEA Board will be supporting these Public Issues Forums for a three year trial period beginning in 2006. The AEA Public Affairs Committee is responsible for identifying the issue, inviting the participants, managing and moderating the forum and disseminating the proceedings. Forums are tape recorded, transcribed and edited into a document posted on the AEA website www.eval.org along with relevant resources and links.

Forum Participants

William Trochim (Moderator) is the President-Elect of the American Evaluation Association (AEA) as of January, 2007 and will serve as President beginning January, 2008. He served from 2005 – 2006 on the AEA Public Affairs Committee and was Chair of the committee in 2006 when this Forum was held. He also served as a member of the Board of Directors of AEA from 1995 to 2000. He is a Professor in the Department of Policy Analysis and Management at Cornell University's School of Human Ecology, and currently serves as Cornell's Director of Evaluation for Extension and Outreach. His research is broadly in the area of applied social research methodology, with an emphasis on program planning and evaluation methods. He is known for his work in quasi-experimental alternatives to randomized experimental designs (especially the regression discontinuity and regression point displacement designs) and for the development of a multivariate form of structured conceptual mapping. He has written or edited numerous books, monographs and journal articles, and is the author of several widely used introductory research methods texts. He is the co-founder of Concept Systems Incorporated, a company that provides software, training and consulting services to support the concept mapping method he developed. Dr. Trochim received his PhD from the Department of Psychology at Northwestern University in the specialty of Methodology and Evaluation Research.

Michael Schooley is chief of the Applied Research and Evaluation Branch, Division for Heart Disease and Stroke Prevention with the Centers for Disease Control and Prevention. In this capacity, he provides leadership and vision to applied research and program evaluation activities to facilitate CDC's response to public health issues related to the prevention and control of heart disease and stroke. Mr. Schooley has been working with CDC since 1993. During this time, he has developed systems for tracking and reporting programmatic indicators, behavioral and environmental measures, and health outcomes to provide data for program planning, improvement, evaluation and accountability. Mr. Schooley has focused on applied research, with particular emphasis on monitoring and evaluating science-based chronic disease prevention and control programs, policies and health outcomes. He has developed performance measures and prepared documentation in response to Healthy People, GPRA, PART, and GAO investigations. Mr. Schooley received his undergraduate degree in psychology from Pennsylvania State University and master of public health in epidemiology and international health from Emory University.

Ted Kniker is an executive consultant with the Federal Consulting Group, a franchise of the Department of the Treasury. He works with federal organizations to improve their performance planning, performance measurement and program evaluation operations and serves as the PART expert for the FCG. Prior to joining the Federal Consulting Group, he was the Chief of Evaluation and Performance Measurement of the U.S. Department of State's Bureau of Educational and Cultural Affairs from 1998 to 2006. In this position, he was responsible for coordinating strategic evaluations and bureau performance planning to ensure compliance with the Government Performance and Results Act (GPRA), the President's Management Agenda, and program improvement. Under his leadership, the evaluation division was recognized as a best practice by the OIG and achieved the highest international affairs score on the OMB Program Assessment Rating Tool (PART). Mr. Kniker received his Master's degree in International Relations from George Washington University and his B.A. in political science from Grinnell College. Mr. Kniker obtained a certification as a Management Planning and Analysis Specialist in 2000, graduated from the U.S. Department of Agriculture Graduate School's Executive Potential Program in 2004, and was appointed in 2004 to the Board of Examiners for the Malcolm Baldrige National Quality Award.

Nancy Kingsbury is Managing Director for Applied Research and Methods at the U.S. Government Accountability Office (formerly the General Accounting Office), where she is responsible for managing GAO's advanced analytic staff including economists, computer engineers, statisticians, social science analysts, program evaluation experts and other scientific specialists. Her team also does a body of work on program evaluation in the federal government, including the implementation of PART. Prior to this appointment, Dr. Kingsbury was an Assistant Comptroller General in the General Government Division (GGD) responsible for GAO's work on government-wide management issues including human capital management and government business operations, tax policy and administration, justice and immigration issues, and financial institutions and markets. She has also served as Director for Planning and Reporting (GGD), Director for Federal Human Resource Management issues, Director for Air Force issues, and Director for Foreign Economic Assistance Issues. She came to GAO in 1984. Between 1979 and 1981, she served as the Director of Resource Management at the Peace Corps, where she was responsible for the agency's personnel activities, budget, and administrative services. From 1972 to 1979, and later from 1981 to 1984, she served in a variety of positions in the Office of Personnel Management (formerly Civil Service Commission), including participation on the team responsible for civil service reform in 1978 and evaluation of the Intergovernmental Personnel Act grant program. Dr. Kingsbury has been active in AEA and one of its predecessor organizations, the Evaluation Research Society, since their inception, and she served many years on the AEA Board and as Secretary Treasurer. She is also a Fellow of the National Academy of Public Administration. She has a M.A. and Ph.D. from The Johns Hopkins University and a B.A. from the University of Miami (Florida).

An Introduction to PART: William Trochim

The topic for this inaugural forum is the Program Assessment Rating Tool or PART, for short, a major initiative of the Office of Management and Budget (OMB) to assess and improve federal program performance, with the goal of achieving better results from such programs. I will introduce the PART process and its website dissemination mechanism, ExpectMore.gov, and describe briefly how they operate. Then each panelist will have a chance to present their experiences with PART and their sense of its strengths and weaknesses, after which we'll take general questions from the audience. Finally, I hope to give each panelist an opportunity for any final summary remarks.

The Program Assessment Rating Tool (PART) can be viewed as evolving out of the Government Performance Results Act (GPRA) of 1993, which mandated the development of a system for assessing performance of all government programs. PART was developed during 2002 and is now in the 5th year of a 5-year cycle to assess all Federal programs. It has assessed 793 programs to date and expects to reach approximately 1,000 in total by the end of 2006. Programs will continue to be assessed at least once every five years.

The PART questionnaire includes 25 questions that are divided into four sections:

- The first section of questions asks whether a program's *purpose* is clear and whether it is well designed to achieve its objectives.
- The second section involves *strategic planning*, and weighs whether the agency establishes valid annual and long-term goals for its programs.
- The third section rates the *management* of an agency's program, including financial oversight and program improvement efforts.
- The fourth section of questions focuses on *results* that programs can report with accuracy and consistency.

The questions in section one, which account for 20% of the final score, ask whether the program's purpose is clear, whether the program addresses a specific and existing problem or issue, whether it duplicates other efforts, is free of major design flaws, and is designed effectively.

Section I. Program Purpose and Design

1.1: Is the program purpose clear?
1.2: Does the program address a specific and existing problem, interest, or need?
1.3: Is the program designed so that it is not redundant or duplicative of any other Federal, State, local or private effort?
1.4: Is the program design free of major flaws that would limit the program's effectiveness or efficiency?
1.5: Is the program design effectively targeted so that resources will address the program's purpose directly and will reach intended beneficiaries?

The questions in section two, accounting for 10% of the final score, focus on strategic planning and ask whether the program has performance measures, targets and timeframes, annual performance measures that can assess progress with baselines and targets, whether partners are committed to working toward these goals, whether budgeting is tied to accomplishments and whether the program has taken steps to correct any deficiencies. I've highlighted question 2.6

which is of particular relevance to evaluators and asks whether "high-quality" independent evaluations are conducted on a regular basis.

Section II. Strategic Planning

2.1: Does the program have a limited number of specific long-term performance measures that focus on outcomes and meaningfully reflect the purpose of the program?
2.2: Does the program have ambitious targets and timeframes for its long-term measures?
2.3: Does the program have a limited number of specific annual performance measures that can demonstrate progress toward achieving the program's long-term goals?
2.4: Does the program have baselines and ambitious targets for its annual measures?
2.5: Do all partners (including grantees, sub-grantees, contractors, cost-sharing partners, and other government partners) commit to and work toward the annual and/or long-term goals of the program?
2.6: 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?
2.7: Are Budget requests explicitly tied to accomplishment of the annual and long-term performance goals, and are the resource needs presented in a complete and transparent manner in the program's budget?
2.8: Has the program taken meaningful steps to correct its strategic planning deficiencies?

The program management questions in section three, which accounts for 20% of the final score, ask whether the agency collects and uses performance information, holds managers and partners accountable, whether funds are handled appropriately, if procedures are in place to measure efficiencies and cost effectiveness, whether the program collaborates and cooperates with others, has strong financial practices, and addresses identified management deficiencies.

Section III. Program Management

3.1: Does the agency regularly collect timely and credible performance information, including information from key program partners, and use it to manage the program and improve performance?
3.2: Are Federal managers and program partners (including grantees, sub-grantees, contractors, cost-sharing partners, and other government partners) held accountable for cost, schedule and performance results?
3.3: Are funds (Federal and partners') obligated in a timely manner, spent for the intended purpose, and accurately reported?
3.4: Does the program have procedures (e.g., competitive sourcing/cost comparisons, IT improvements, appropriate incentives) to measure and achieve efficiencies and cost effectiveness in program execution?
3.5: Does the program collaborate and coordinate effectively with related programs?
3.6: Does the program use strong financial management practices?
3.7: Has the program taken meaningful steps to address its management deficiencies?

The results and accountability questions in section four, which accounts for 50% of the overall rating ask whether the program demonstrates progress toward goals, achieves annual performance goals, shows improved efficiencies and compares favorably to other programs. Of special note to evaluators, the final question explicitly asks whether independent evaluations of sufficient scope and quality indicate that the program is effective and achieving results.

Section IV. Program Results/Accountability

4.1: Has the program demonstrated adequate progress in achieving its long-term performance goals?
4.2: Does the program (including program partners) achieve its annual performance goals?
4.3: Does the program demonstrate improved efficiencies or cost effectiveness in achieving program goals each year?
4.4: Does the performance of this program compare favorably to other programs, including government, private, etc., with similar purpose and goals?
4.5: Do independent evaluations of sufficient scope and quality indicate that the program is effective and achieving results?

In the simplest case, each of the questions is answered with a Yes or No answer and questions within sections receive equal weight, with a total of 100 points for each section. The section scores are then weighted to achieve the final score, which determines which of the four overall categories the program will be assigned to: *Effective* (if the score is 85-100); *Moderately Effective* (70-84); *Adequate* (50-69); or *Ineffective* (below 50). Regardless of overall score, a fifth possible rating of *Results Not Demonstrated* is given when programs do not have acceptable long-term *and* annual performance measures. A program also gets a rating of Results Not Demonstrated when it lacks baselines and performance data to indicate how it has been performing.

The current distribution for the 793 programs that have already been through a PART review shows that approximately a quarter of them received a rating of Results Not Demonstrated. Fifteen percent were determined to be effective. Approximately 30% were found to be Moderately Effective or Adequate. Only 4% have been determined to be Ineffective. Every part review includes an improvement plan that the program can follow to address identified performance issues.

PART: Results to Date

Rating Percent (and Number) of Programs (N = 793)

Effective	15% (124)
Moderately Effective	29% (231)
Adequate	28% (219)
Ineffective	4% (28)
Results Not Demonstrated	24% (191)

The website ExpectMore.gov, shown in Figure 1, was launched in February 2006 and is the primary public mechanism for reporting on Federal program performance and what is being done to improve results. The site includes a variety of simple mechanisms that enable the user to identify programs that are performing or not performing, and search by keyword or topic.

Figure 1. Main page from ExpectMore.gov, the web site displaying results from PART assessments of Federal government programs.

As one example, Figure 2 shows a section of the report displaying programs that are considered to be performing, in the section of reviewed programs from the Department of Education.

Figure 2. Summaries of PART assessment results for programs in the Department of Education from ExpectMore.gov.

ExpectMore.gov is designed to be a user-friendly public site, which provides simple explanations for the ratings and uses different numbers of stars to indicate the final rating category for each program. The user can then drill down on that overall rating to get more detail about how that assessment was reached and what the improvement plan is for the program. For example, Figure 3 shows an example of the rating details for the Even Start program, which was judged to be ineffective.

Figure 3. Rating details for the Even Start program.

As would be expected with any federal program of such magnitude, the PART process and the ExpectMore.gov website have not been without controversy. Several reports of the Government Accountability Office have raised concerns about implementation and advocacy groups outside the federal government, such as OMB Watch, remain critical. Within the evaluation profession, a lively discussion has been joined about the degree to which PART reflects good evaluation practice, and that discussion is one reason we decided to focus on this as both a timely and important issue for this conference.

There's an old saying that the "devil is in the details" and that is certainly likely to be the case with a system that is as ambitious and complex as PART. It is impossible in so short an introduction to do justice to the many details and controversies that swirl around each of the specific ratings, let alone to consider the broader implications of their aggregation, categorization and representation to the public. And so, we present this first AEA Public Issues Forum in the hope that it will begin to identify some of those issues and present the experiences and reflections of evaluators who

have been involved in different ways with the PART system, which has such important potential implications for the theory and practice of evaluation.

Panel Comments: Michael Schooley

Good afternoon. I want to start by recognizing and thanking Bill for his introduction and warm remarks, and also for organizing this panel. As he mentioned, there was a lot of discussion on EvalTalk earlier this year on two different occasions, so I think we all welcome the opportunity to engage in this fruitful discussion.

I'm going to start by adding to a few of the comments Bill has made and reiterating a few of the points that he highlighted in his introductory remarks. I'm going to talk about some of the PART reviews that have been conducted at CDC, how they have been conducted and how the agency has done on these reviews, and then I'm going to share some of my own observations and considerations as an evaluator working on some of the reviews. If you've heard a few other federal presentations, you'll not be surprised to hear me say that the views I present are my own and do not necessarily represent those of the agency.

As Bill mentioned, PART was created in 2002 – fiscal year 2004 was when it was first applied, to programs at CDC and elsewhere. It was presented as a diagnostic tool to be used by OMB to assess program performance, and its intention is to improve program performance.

The tool, as Bill introduced, is 25 to 30 questions, divided up into four broad sections or areas. One of the questions of particular importance for evaluation is, "Does the program have a limited number of specific, ambitious, long-term performance goals that focus on outcomes and meaningfully reflect the purpose of the program?", "Does the program achieve its annual performance goals?", and then "Do independent and quality evaluations of the program indicate that it is effective in achieving results?"

The PART tool can be slightly adapted across a range of programs. There are seven broad areas of programs it can be applied to:

- Competitive grant programs
- Block/formula grant programs
- Regulatory-based programs
- Capital assets and service acquisition programs
- Credit Programs
- Research and development programs
- Direct federal programs

Most of the work that is done at CDC falls under the competitive grants program. At CDC 23 programs have been reviewed over the past five years, and we are in the final year of evaluating programs for fiscal year 2008. Fifteen out of these 23 programs are competitive grant programs, and the funding range is anywhere from $63M to $900M. I think this is important in illustrating that there is sufficient variation in the size and scope of programs that have been reviewed, even within CDC and the public health realm.

Here are examples, in terms of content and topics, of some of the programs that have been reviewed at CDC. These include things like the child immunization program, and in comparison, at the end of the list, the global immunization program, as well as an entire agency, the Agency for Toxic Substances and Disease Registry. Other areas include state and local preparedness grants, domestic HIV/AIDS program, chronic disease prevention and health promotion programs,

which represents a variety of programs addressing tobacco, cancer, heart disease, nutrition, and physical activity.

Sample CDC programs that have had PART reviews

- Childhood Immunization Program
- Domestic HIV/AIDS Prevention
- State and Local Preparedness
- Chronic Disease Prevention and Health Promotion
- Agency for Toxic Substances and Disease Registry
- Global Immunization

In terms of the reviews that have been conducted at CDC, most of the reviews have resulted in scores that are adequate to moderately effective. So overall, as an agency, we've actually done pretty well if you're just looking at the final scores. One program was designated as effective, five programs were determined "results not demonstrated", and we haven't had any programs that have been determined to be ineffective.

In general, if you look across those four broad sections, the lowest scores are on the results sections – that's where I think a lot of challenges come in with programs in providing and demonstrating results and providing some of the evaluation information. Over the past five years, the scores have increased, overall by about 10 percent. This is actually interesting, and I think Bill made an allusion to this when he said, "Ted obviously knows how to do a good PART review" - in that the agency, in some respects, may have become a better test-taker, in knowing what to provide for the review and how to prepare for the review. The other issue is that over the course of the five years we have had three OMB examiners. We had one OMB examiner for the first couple of years of reviews, another one for the second couple of years, and just recently a third examiner for the most recent year. As OMB examiners change, there is some subjectivity to the review and the exam, and that may influence the review and the scores that we receive.

And then finally, part of the PART review is looking at performance measures. These measures are generally outcome measures, mostly longer-term health outcome measures – not shorter-term measures, and not really a lot of output measures. Some samples of these measures at CDC are listed here.

Sample performance measures at CDC

- Number of countries in the world with endemic wild polio virus
- For states receiving CDC funding, increase the percentage of persons with diabetes who receive annual eye and foot exams
- Reduce the prevalence of Chlamydia among high risk women under age 25 by 15%
- Ratio of operational costs to total program costs in countries receiving $1M+ HIV/AIDS

There's a document on the guidance for OMB website, which documents the workshop that occurred in 2003 looking at performance measurement issues in providing measures for PART. It addresses seven broad areas around performance measurement challenges. I think these are of particular interest to program evaluators, and obviously ring true to a lot of the work we do, especially public health. As programs have been preparing for these PART reviews and preparing performance measures, some of the things that come up are:

(1) Program outcomes are extremely difficult to measure. Some of the programs are looking at broad and far-reaching things like policy change and issues, which are difficult things

to measure from a scientific perspective and from a specific discrete measurement perspective.

(2) The programs oftentimes are one contributor to the outcome, looking at the issue of contribution versus attribution.

(3) The results often will not be achieved for many years. There is a long time lag between the program implementation and the actual results.

(4) The programs, especially in public health, often relate to deterrence or prevention of specific behaviors and not actually seeing a change or reduction in specific behaviors, which can be another measurement challenge.

(5) Oftentimes programs have multiple purposes and funding can be used for a range of activities.

(6) Looking across a variety of sites, you may see a range of things going on that are not necessarily a standard program being implemented in every site.

(7) Sometimes the program purpose is administrative or process-oriented – one unique thing with many of our programs, a lot of times the leadership, management, and administration costs are a separate line item within the budget and might be looked at separately – how do you look at the outcomes for those types of things outside of the program context?

In addition to performance measures, there is one other key outcome of a PART review, and that is the program improvement plan. That is one of the things that actually comes out of doing a review, so I wanted to share, just in terms of impact on implementation, two very brief stories about program improvement plans. These highlight what we've seen with some CDC programs and demonstrate some good use and application of the PART reviews.

- In immunization, one of the things that came out of the PART review, and going through the PART process, was an improvement plan to actually integrate and automate vaccine ordering and management by centralizing the distribution of all public-purchased vaccines, to try and get at better efficiency and delivery of vaccines to the public.

- Within the Agency for Toxic Substances and Disease Registry, one of the things that came out of doing the PART review and assessment was a renewed commitment to evaluating the interventions by documenting effectiveness at each hazardous site to determine its particular and specific impact on public health.

There has been a lot of discussion and a lot of concern about budget implications, and how PART might be used as an "axe" or a tool to put programs on the chopping block. I can only speak to what I've seen and experienced at CDC – I know there are a wide range of other stories out there about programs, but I'll share with you some of my observations at CDC.

OMB does consider PART in developing the President's budget. So looking at the PART review and the PART scores is one of the tools that OMB uses in the preparation of the President's budget. However, the scores are generally not considered and do not come up in Congressional review of the budget. So, things that may go into OMB's preparation of the President's budget could get reinstated during Congressional review without regard for what the PART score or the PART review was. That is just one example, but generally we haven't seen a lot of use or value for the PART scores in the Congressional review process.

In looking at just a snapshot three-year period – and I didn't do this statistically, I just eyeballed it – but looking at a three-year period of PART reviews and funding cycles for programs, I really can see no clear association, positive or negative, between PART scores and funding changes in programs. There were a few that went up, a few that went down, none of them significantly or substantially. I did mention earlier that CDC hasn't had any programs that have been

demonstrated ineffective, so the way our agency has done and our scoring may have contributed to that, but there has been no clear association that I could find. That said, as budget constraints continue to increase, and funds become tighter and tighter, I don't think we can say that there isn't great potential for PART to be increasingly used in budget decision-making.

In thinking about implications of the PART process, and the PART review, there was a little aside in an on-line forum that I found and just couldn't help but share. It was conducted in 2003 with Clay Johnson, then the deputy director for management at OMB, and the question was, "What is the penalty for an agency if they get a low score?" And his response was "Public shame and humiliation, and the opportunity to be questioned about it by the President." Which I think is summed up nicely in a cartoon of the poor lone government official defending his program, where he is surrounded by cameras and microphones saying "Accountability (stinks)".

So shifting towards the close of my presentation, here are some observations and considerations:

• One of the key aspects of any PART review is the OMB examiner. Their investment in the process and learning about the program is critical to what comes out of the review. Just as in evaluation, the investment and involvement of the program evaluator, in working with the program, is critical to what comes out of that evaluation or that assessment.

• Consistent interpretation of the guidance is critical. PART is designed to be an objectively applied tool and process, but there is surely subjectivity – whatever the adage is, "interpretation is 9/10ths of the law". There is some variation in interpretation, and that is one of the things that the OMB guidance is trying to address, that variation in interpretation of the rules.

• Finally, with respect to the examiner, knowing what is acceptable is important – having clear communication with the examiner of what they really want, what they are trying to get at, and trying to understand what they need, so that they can answer the question and score the question appropriately.

On the other side, involvement and investment of program leadership in the process is critical. PART can be an extremely resource-intensive process to go through, so having leadership engaged, involved and invested in the process – to not only conduct the review, but actually do some follow-up after the review – is of critical importance. So the leadership is very important to the process and what comes out of the review.

In general, I think that PART overall has been a continuing boon to evaluation in Federal government, in that it clearly states and demonstrates that programs should be evaluated. That's not a bad thing from a program evaluator's perspective. Because of the questions, and the way that program evaluation is displayed in PART, it has elevated the importance and attention to evaluation. Now sometimes, not all attention is good attention, or you could argue that any attention is good attention – but the importance, the commitment to evaluation within the agency, I think has continually increased, and PART is one thing that has contributed to that, especially in recent years.

During a review, there also have been dedicated resources for evaluation. When people are mobilizing for these reviews, there are often people who are pulled off of their regular jobs and dedicated to nothing but the PART review process; usually for at least six months, sometimes for much longer than a six-month period. There are workgroups that are called together, and program evaluators are pulled in and made part of that process. So there are at least, during that time, some dedicated resources for evaluation.

In general, if you look through the tools and the framework, the approach, intent and questions of the PART review, I really think that they are helpful and informative. From an evaluator's perspective, and in applying an evaluative process to programs, it has merit and some good things that can be applied. And then, in terms of implementing it within the agency, it provides the opportunity to engage multiple internal stakeholders. When doing a PART review, there is increased attention to the program, there is a substantial opportunity for dialog among program planners, decision makers, evaluators, all coming to the table talking about issues that even in their day-to-day work they don't normally take the time to think about or talk about.

On the other side, some of the challenges with respect to the PART review process are as follows:

• PART is really designed, in some respects, as a one size fits all approach. There are slight adaptations that can be made across programs, as I mentioned earlier, but it really is one tool to fit the breadth of federal programs, and there is great variation in program size and scope.

• The resources invested in evaluation and reviewing programs is mobilized during the review, but afterwards, I think there are questions and concerns about the sustainability of those resources. So when the dust settles, and everyone goes back to their regular jobs, you might forget about all of those great things you learned while going through the process and reviewing the program.

• In thinking about the ExpectMore.gov website, there are issues surrounding the reporting of findings. Even though you can drill down, a lot of people probably don't drill down, so you just get the primarily quantitative score for the summative report, which is probably one page or less – a brief summary of the program. From a decision-maker's standpoint or a legislator's standpoint, there is certainly concern about understanding that there is a lot more behind that evaluation and what you need to know about that program in order to make decisions about it.

• There has been some historical debate on evaluation evidence and realistic outcomes. While I am not going to touch on this in detail, there has been a lot of work over recent years to increase the breadth of information in evaluations that will be accepted as evidence for program evaluation, not just basing things on a medical research model paradigm, and only accepting a higher gold standard of evaluation.

• And then finally, with respect to challenges, there is often an approach, especially from program planners or decision-makers, that performance measurement = evaluation = program improvement – if you do any one of those things, or address any one of them, you are addressing all of them. And that is obviously not the case. Performance measurement is an important tool, but that does not absolve you from doing program evaluation, and either of those does not necessarily relate to anything being used and actually seeing improvement in the program.

So in closing, I think one of my critical issues is that PART has a lot of good aspects and good components. Ensuring that there is a continued organizational commitment to evaluation, and actually institutionalization of those evaluative processes that PART brings to programs in Federal agencies, is where we have the next challenge, and where we can move forward with things. With the PART review, a lot of people focus on the destination of getting to the score, and getting the review done, but there is equal value, I think, in the journey. We need to think about how to use our learning from the journey in our continued work, and as we move forward. And then there are a lot of questions with respect to the future of PART. I don't have the answers, any

more than anyone else at this table does, but that is something for consideration - is this is going to be around in the next administration and the coming years, or is it not?

I will close by acknowledging and thanking some colleagues at CDC who helped prepare for this, including Julie Zajac, Marcia Taylor, and Thomas Chapel, as well as mention that there is a website available if you want more information on PART:

http://www.whitehouse.gov/omb/part/index.html.

Thank you.

Panel Comments: Ted Kniker

Good afternoon. It is quite an honor for me to be here for the first AEA Public Issues Forum, and I am humbled to share the dais with these three distinguished people. I feel like a rookie who has been called up to the majors. But I appear before you today as a survivor of the PART, having been personally responsible for three separate assessments, advising on numerous others, and now serving as a consultant to other Federal agencies in how to improve their performance and build evaluation capacity. And, of course, also being a "Fed", I must have those disclaimers – so the views expressed in this presentation are my own, and do not represent the official views of the Department of State, the Department of the Treasury, the American Evaluation Association, or CBS and its affiliates (laughter).

Now, we've all heard the phrase that where you stand depends on where you sit, and I think that this is certainly true from my perspective. As a program manager, and the chief for evaluation in public diplomacy for the Department of State, I was first very excited about PART. But after the first year, I thought that PART had a different meaning. I thought it stood for "Pernicious Annual Required Torture". (laughter) I think it's important to note here that PART is a laborious process, both for the agencies and the OMB examiners. I think we tend to forget that OMB examiners who work on PART are generally working on three to six others at the same time – so, they are not always pleased with this process either.

However, as the Bureau of Educational and Cultural Affairs gained success in the PART, and now as I've transitioned into my role as a consultant, I think I've kind of changed my view on PART to some degree, and now, it has a new meaning for me, which is "Program Analyst's Retirement Trust". (laughter) But in all seriousness, though, I have always considered myself a supporter of the PART, because while I've been an advocate of the PART, I also acknowledge that it is not perfect, and it does need to be improved. But where I've been a supporter of it is as a management tool -- a management checklist -- not necessarily an evaluation tool.

The first thing that anyone should know who lives outside of the Beltway is that PART has become so ingrained in the Federal culture it is now a verb. A little history that you may not get from the general readings that you do or from OMB is that in the fiscal year 2002 budget, the President's budget, there were programs that were given ratings. The program that I was evaluating at the time was given a "moderately effective". Now this came before anyone had ever heard of PART, and it just showed up in the budget, and so there was a lot of uproar from several agencies about "how did you make this assessment?", because it wasn't transparent at all to the agencies, and OMB came back and said, "Well, we kind of did this internal thing, and we are thinking about maybe doing some kind of assessment in the future, and we just piloted this, but we didn't tell anyone about it." So the next year, in FY 2003, there were some pilots, and I helped to work on one of those with the Department of State on the Freedom Support Act, which was technical assistance provided to the former Soviet Union for democratic and market reform.

Now, I'm going to talk to you mostly from my experience at the Department of State, which I only left about six months ago. And for those of you who are unfamiliar with the U.S. Department of State, its mission is to create a more secure, democratic, and prosperous world for the benefit of the American people and international community. And the Bureau of Educational and Cultural Affairs' mission is to improve and strengthen the international relations of the United States by promoting mutual understanding between the people of the United States and other countries – a very observable and measurable thing to deal with.

The Bureau, for those of you who are unfamiliar, exchanges about 35,000 people per year through 90-plus programs. Some of these programs you may be familiar with, such as the Fulbright Scholarship program, the International Visitor Leader program. The programs range from high school to senior government officials, business leaders, and administrators at universities, and they are administered by over 100 non-profit partner organizations and 200 colleges per year. And it's done with a budget of approximately 355 million dollars, of which the amount that went to program evaluation was .43 per cent, or about 1.5 million dollars. Now, over time, that went up a little bit, but our budget also went up a lot, so the actual percentages remained constant.

Our PART story

When we started, and when we were told that PART was coming, we were really excited about it. And the reason that we were was that we were the only bureau in the entire Department of State that a staffed and resourced internal evaluation unit. So we thought we had this thing nailed. We said, "we've got the evidence of effectiveness, we have everything, there can be no problems." And when we got our first report back, our first score was a 61% "results not demonstrated", which was kind of a shock to us. But the reason that OMB gave to us for why we had "results not demonstrated" was because we had performance measures that were not linked to long-term goals, that did not have baselines or ambitious targets, and we had an unclear strategic planning process.

So in the next year, fiscal year 2005, we made some revisions. We refine our measures, we hired a performance measurement expert to come in and really work with us to re-align our measures with our strategic plan, to help us actually draft measures that read well. We linked PART to planning, and this was also something that was being done in the Department at the time. There is a report that has come out recently by John Gilmour from the College of William and Mary that talks about integrating budget and performance, and one of the examples he uses is the Department of State, and how we had a centralized resource management office that took PART very seriously and very aggressively, and worked to ensure consistency within the organization and to build PART within our strategic planning process.

In the bureau we also implemented frequent measurement. Now one of the things that is important to point out here is while it looks like there is a year between all of these, the actual time we had between assessments and re-assessments is actually about six months. The process that OMB goes through works so that you know your first inkling of your score sometime in the late summer. Generally by October you have a good idea of what it is, and then in March again, they are in there re-assessing. So you have from basically October to March before you really know what it is they found wrong with you, and it's written down to a point that you can take action on it.

Well, then in fiscal year 2006, as we had done all these things and continued to refine what we were doing, we ended up with a 98% effective for that first PART, and then we also had a second PART that was launched, and we came out of the gates with a 97% effective rating.

FY 2004 →	FY 2005 →	FY 2006
61% - Results Not Demonstrated	92% - Effective	98% - Effective
		97% - Effective
Measures are not linked to long-term goals, do not have baselines or targets	Refined measures	Refined targets further
	Hired performance measurement expert	Used measurement data for budget request
Unclear strategic planning process	Linked PART to planning	Implemented systematic measurement
	Implemented frequent measurement	

Figure 1. Evolution of PART assessment ratings for the Bureau of Educational and Cultural Affairs in the Department of State from fiscal years 2004 through 2006.

The results of PART

Now, it's important to remember, I think, that PART is as much a process and a relationship as it is a tool, and a significant element of that process and that relationship is communication. And I always used to say that PART is not about compliance, it's about alliance. It's about figuring out how to get yourselves aligned with OMB, in order to ensure that they understand what you're doing and can be an advocate in the budget negotiations for you. Better communication with OMB was probably our first result. Prior to PART, the only folks in our bureau who talked to OMB were our budget people. After that, our leadership and our evaluation people had regular contact with OMB. I probably spoke with our examiner at least once every couple of weeks, and we were constantly feeding information about evaluation and what we were doing.

We also had internal cultural change, and we moved from kind of your normal government processing – "move money out the door, make sure you spend all of your money before the end of the fiscal year" – to really a focus on results, and what it is we are trying to achieve. And I think that more than GPRA or the President's Management Agenda, PART prompted real discussions in our senior leadership meetings and among staff about what those results were – and what they should be.

We also improved planning and focus of our programs. The bureau's primary business line is grants, and you have ever read RFPs, basically from the 1990s all the way through to a few years ago, most of those RFPs for grants would focus on the activity that they wanted the recipient to work on – "you will implement an orientation program, you will do this, you will do that". And it wasn't until a couple of years ago that we started talking in those RFPs about what it is that we are trying to achieve, and what it is we want our organizations to work towards with us. So we think we had better grants after we created a stronger culture of measurement.

We also re-directed resources towards performance measurement and away from evaluation, and this is something that always bothered me. But what happened was performance measurement became so important in the PART that that collection and analyzing of data, we had to put resources into it, and because resources weren't necessarily going up in a way that we

would have liked, we had to take what we had and to re-allocate it. We doubled the staff in our evaluation unit. We originally had four people working on evaluation, and it became basically nine people. We worked back to three people doing program evaluation, and the rest working on systematic program performance measurement.

We also increased funding for programs in evaluation. Now, that may not jibe with what I just said, but eventually over time, because we had done so well on PART, it was something that we used in our budget negotiations, and it was something that was held up within the department as, "here's the example of a bureau that is doing it right, doing it well, and they should be rewarded". Now, I won't say that there is a causal link there, but it was certainly correlational. There were some other things that were happening in the environment that allowed exchange programs to gain more and a bigger share of the budget, but still there was the evidence that those kinds of programs worked.

We integrated our planning and evaluation better – and again, this is what I was talking about earlier – that in our Congressional budget submissions, we had all of our PART information in there, we had all of our measures. We spent a lot of time talking about PART in Congressional budget submissions.

We also had more respect and recognition. I think that one of the important things PART has done is make government programs transparent. It has helped to really shine a light on what we're doing with your money. We were held up as a model, and that was really great for us. We were benchmarked by a number of different agencies and firms, and we had consultants coming in and talking to us about how we did our work, and of course then what success leads to is more work. We ended up being asked to do a lot more, as our audience and our constituents became more sophisticated, because we had really outreached to them to educate them about what we were doing. They started asking more sophisticated questions, which led to, "Oh boy, now we've got to start measuring this even more rigorously".

Challenges with PART

Of course, success doesn't come without a price. If there is no pain, there is no gain. So we had some challenges. And one thing I'll state here is that PART is not coordinated or accepted by Congress, plain and simple. We had a particular incident when I was called to the Hill to brief a staffer on our appropriations staff about why public diplomacy in general didn't have measures, and when I said to her that, "What we can do is show you our PART measures, and we'll report those on a quarterly basis like we do to OMB", the response was, "The PART is OMB's. It is not Congress's. I am not interested in looking at what you submit to OMB." When we pressed a little bit further about what information they wanted, that was different from what we were submitting, we didn't really get a response – we basically got, "Well, you figure it out." And then when we kind of thought we had figured it out, and started sending it around, what it ended up turning out to was that it had nothing to do with results. It was, "We just want to know what you spent the money on, and what you got out of it in terms of output" – it really simply boiled down to output.

The Paperwork Reduction Act requirements, versus reporting results requirements, one of the interesting things we encountered was, as the PART kept moving, all of a sudden you noticed that you were reporting more frequently. It went from annual to quarterly. And so we had all of these quarterly management reports we had to submit, and that was coming from the budget side of OMB – yet, on the regulation side of OMB, there is the Paperwork Reduction Act, which requires that a survey of ten or more individuals conducted on behalf of or sponsored by a Federal agency must have an OMB approval. The OMB approval process takes, in our estimation and experience, six to eight months. And then on top of it, they weren't granting generic

clearances. So we had no way of getting clearances for our surveys in a timely way, in order to collect the data to give to the other side of OMB.

We saw that in the improvement plans, they went from constructive to idiosyncratic. One of the benefits is that as your OMB examiner becomes more familiar with your program over time, they in my experience typically become more of a supporter of it. But we have had three OMB examiners in four years, and so that became kind of an iffy process. But we noticed that if they stayed around too long, what would start happening is that they would tell you that how to improve was some wild idea. So we already had this 97 and 98 per cent effective ratings on our two PARTs, and all of a sudden we got lumped in to a couple of other PARTs that were being done for the Department, and then all of those together were collectively told, "You won't succeed unless you have an over-arching strategic plan" for that particular part of the Department.

As for the role of evaluation, it was very clear to us that it was more important to focus on performance indicators than on evaluation. That first year, when we had the 61 per cent, we actually had 20 separate, independent evaluations that demonstrated effectiveness across a range of our programs that we handed in as part of our evidence, and it didn't really count. We ended up getting, I think, a small extent rating for that. So it was very clear that either there wasn't an understanding of evaluation, or there wasn't an appreciation for the results that were demonstrated in an evaluation. It was all about the measures, and the indicators.

And then there were some issues with the definitions of independence and quality. We've definitely seen this across the board – in some cases independence is only seen as possibly a GAO or an OIG audit, because those are not funded. We had problems with one of our examiners that, even though we were an internal unit that was separate from program management, and actually still funded external evaluations, that because we funded them, that was a problem.

There was also, we found, little room for qualitative data or citizen satisfaction. One of the things that I've noticed is that in the Clinton-Gore team, citizen satisfaction was really the main thing that the administration was promoting – that it was important that citizens be satisfied with government services, and what government was doing. And that has kind of moved to more of what I view as, if you liken it to business, a move from customer focus to shareholder dividends. The focus is improving the dividend, as opposed to really whether the beneficiaries are supported and happy with what you are doing.

Also, OMB also – and I'm glad that we don't have a representative here, so now I can say this – is shifting boundaries and the underlying philosophy of PART. PART has a philosophy that no program is perfect, and that you have to show improvement. And we had an issue in which a lot of our measurements showed that we were about as high as we were going to get with things. One of my bosses called it "the Lance Armstrong effect" – if you are Lance Armstrong, and you've won the Tour de France six or seven times, what do you do next to improve? You could win it in a few more minutes, or you could maybe win it one more time, but at that point who cares? It really doesn't matter. And so there is not a focus from OMB on maintenance of excellence. We also found that there were requirements added after the fact, and I mentioned a couple of those.

Observations

I just wanted to make some observations that I've seen from other agencies, to wrap up really quickly:

- PART is seen as compliance, instead of a way to communicate. Agencies see PART as increasing their bureaucracy or taking away from their ability to program.

- They struggle with output versus outcome.
- They are confused about the role, definition, and importance of evaluation.
- Agencies don't want to be held accountable for societal impact, because they feel that if they are held accountable for it through the PART, they will always lose.
- The PART rating, as was mentioned earlier, generally is dependent upon an individual examiner, although there are some processes in place at OMB to change that.

My assessment of PART

Here is my take on a quote from Shakespeare: "They say he PARTed well , and paid his score; And so, God be with him!" I used this quote with my examiner to say, "If Shakespeare said it, and we did it, we should get something good," and then she reminded me that this quote comes from Macbeth, which is about death and destruction. (laughter)

So, does PART improve program performance? My assessment is that results are not demonstrated. It does, however, improve goal articulation, planning, and stakeholder dialog, as well as elements of program management. It is kind of mixed as to whether it actually improves performance – one of the things you'll see is there is no consistency on what is accepted as the measure. Some agencies are given certain measures and are told, "these are great", while another agency will use a very similar measure and are told, "no, that is not acceptable".

There are also issues where it does generate attention to performance measurement. Evaluation certainly increased internally, and our reputation as well as respect for what we did in the Department of State certainly increased. And there certainly now seems to be evidence that evaluation is important.

And does PART need improvement to foster the improvement that it claims to achieve? I think we need to change from a "yes/no" format to a continuum – that if they allowed a continuum of "this is where an agency is", it would be much easier to build agencies towards improvement by providing agencies with a "ladder" of success to follow.

PART does not include questions for assessing leadership or workforce development, two things that I think are critical for an effective organization. And it doesn't currently allow for the contribution of programs – there still is a sense that it has to be a causation.

As for the future, I think that PART, or something like it, will continue to exist. We have gotten too far into results measurement to move away from it – in fact, over the last couple of years, the Malcolm Baldridge Quality Awards have now opened up to non-profits and the government. And ExpectMore.gov is out there now – people are reading it, people see it. And so I think there will be a continued need for these kinds of things. Thank you.

Panel Comments: Nancy Kingsbury

I had some trepidation in following Ted Kniker – he has done presentations for my classes, so I know what he is up to, and it is going to be a hard slog to be that much more entertaining.

My role in this panel, I think, is to step back and put some context around some of these issues. PART has, in some respects, come further than I would have expected it to when it first started, and I think it has some good things to say for itself. But we start at GAO from where our current leader, the Controller General of the United States, likes to start every single talk, which is that the United States government is on a fiscal path that is fundamentally unsustainable into the future, and we need to develop tools and techniques, and most importantly the political consensus to make changes in government programs so that we can avoid leaving our children and grandchildren in dire straits. That is kind of where we start – and so philosophically, the idea of being concerned about performance, making decisions with information, doing evaluations, and all of that sort of thing, fundamentally has some real value. Whether or not this particular tool gets us there – and I personally think not – it is certainly something we can start from.

Let me go back to something that has been referred to both in this session and the previous session, which is PART as an outgrowth of the Government Performance and Results Act. GAO was there in the development of GPRA. We were very supportive of it, and it had some real features that we thought were important. One was that it specifically called for evaluation to be done. It called for strategic planning and performance planning to be done, and in the context of strategic planning and performance planning, it explicitly called for the results of evaluations to be reported and used. So the framework has always been there.

Now GPRA had its downsides too. Necessarily, it was agency-focused. Arguably, a lot of the decisions that have to be made cross agency lines, and PART certainly has some of that same characteristic. GPRA called for a government-wide strategic and performance plan, which has never happened. One would hope, if there were a government-wide plan, we could begin to get at some of the interaction effects among agencies, and so forth. The President, by the way, and OMB have argued through two administrations that the President's budget is the government-wide strategic plan, but that is not very persuasive.

I bring to your attention that GPRA was passed at a time when the administration was in Democratic hands, and at least part of the Congress that passed the law was in Republican hands. It is interesting to me that GPRA, to its credit, specifically called for the input from stakeholders, broadly defined, in the development of strategic and performance plans – stakeholders to explicitly include, among others, program participants, state and local governments who are affected by these programs, and the Congress. And we did some early reporting on how well that happened. In the programs where it happened well, I think you moved a little closer to getting a consensus about what performance measures should be used, and how to measure the results of programs.

You will recall, however, that the Clinton administration's principal initiative was something called "Reinventing Government", which in many respects was not particularly based on evidence and performance. So GPRA erupted at kind of an interesting point in time. Along comes the Bush administration, who adopts concern about results and adopts a management agenda – a huge step forward, in our view – and then develops this PART. And Ted was right, PART was developed kind of in the back rooms of OMB in its first iteration, and there was an attempt on the part of the OMB management to get GAO to go on record saying that we thought this was the

greatest thing since sliced bread. We refused, tactfully, to do that at the time, but we did provide some input about issues such as whether yes/no questions are really a good idea, and things of that sort.

But PART's unit of analysis is the *program*. Now one of the things I think is not as evident to the evaluation community as might be the case is that "program" is defined as the line item in the budget. Sometimes that has correlation with what you and I as program evaluators would consider to be a program, but many times it does not. One of my favorite examples in that regard is that all of the tactical aircraft programs of all four services in the military are considered a program, subject to PART. This makes no sense at all to me, and I never have figured out how they did that.

If a program can be identified and is coherent, then you apply these questions, but even then, it is a two way dialogue. It is the agency and OMB, with OMB being the elephant in the room. So what happened in many programs' experience is that the performance measures that eventually got settled on were not performance measures, which under GPRA might have included the views of the state and local governments and the program participants or Congress about how to measure performance. They were, rather, ones that the OMB budget examiner liked or wanted to see. OMB budget examiners are, by and large, not trained evaluators. So it was a very ad-hoc, subjective process, and I think has largely remained so. All of the things that have been said from the agency perspective about the value of some of this process – calling attention to performance measurement, calling attention to evaluation, modestly increasing the actual investment in evaluation, I think are all things to the good – but getting there, I think, has arguably been painful for a lot of people, and taken up a lot of management resources in the government.

As the PART process got started, and as an outgrowth of some work we had done under GPRA, we began to re-recognize, because we had been through this some years before, that there was some evaluation going on in the Federal government. It was out there, but it was pretty decentralized and pretty hidden. This effort takes off from the fact that GPRA does require evaluation and so we did a series of engagement that looked at the question of what were agencies doing in the way of evaluation, and how was that helping to frame how the thought about performance measurement, and how they thought indeed, in due time, about the PART process.

Just for our own information, we began to assemble more or less an inventory of where evaluation was being done, and from that we established something that has come to be called the Federal Evaluators Network. GAO has sort of had some nominal leadership here, but leadership has also come out of the agencies. And that network actually began meeting before the PART process started, and so by the time the PART process started and in particular when the debate erupted about methodology and independence, a group existed that could become a focus to discuss what was going on. In that discussion, both issues were equally important.

We had a forum to begin to debate these things among the evaluator community inside the Federal government, and I think that forum has actually proved to be quite useful in helping to mediate some of the discussion about those issues. For example, OMB put up on its website and started telling agencies that no evaluation was worth anything unless it was done by GAO or the IG's. That was just nonsense. And we worked overtime, through the Federal Evaluators Network, to develop a thought process that says, "No – excuse me – the value of an evaluation is whether or not the researchable questions and the methods you are using are suitable to the answers you are looking to get, and that just because it's funded by the agency, so, you still can independently judge the methodologies. That was a debate which we were able to begin to have.

Regarding the other question about the only acceptable methodology being the randomized controlled trial, I think at least inside the government, the Federal Evaluators Network provided a vehicle to begin to push back at OMB on the issue. Eventually, the group came together and developed a briefing – I was going to call it training, but one doesn't want to say that word in front of OMB examiners, they take offense – so we developed some briefings, which we have now done two series of over the last couple of years, to get OMB examiners to better understand what evaluation is, what evaluation methods are appropriate under which circumstances for what programs, and in what phase of development those programs may be. Whether we have actually influenced the decisions that have been made, I'm not sure, but it has gotten sufficient acceptance that some of this material is now actually found not only on the Federal Evaluators website, which is housed by the Environmental Protection Agency, but also on OMB's website itself. So I think we've made some progress there.

But if I could return for a moment to the idea that we really need a process sort of like this, we need it to be much more broadly grounded. It needs to include the Congress. When we went up for one of our engagements to talk with people in the Congress about what they thought of PART, a substantial number had never heard of it, and the ones that had didn't like it very much, because they were never consulted. And it is a reality in our political world that there is a fundamental constitutional issue of the need to involve the Congress in these kinds of decisions that this administration, for whatever reason, has chosen not to do – about a whole lot of things, by the way, and not just PART. Mind you, all this debate about PART and Congress not liking it took place in a situation where the administration was Republican and so was the Congress. This administration has chosen to exert its executive authority in a way that reduces Congress' authority, or attempts to, and Congress quite frankly, Republican or Democrat, doesn't like that very much.

So if we are going to come back to this fiscal sustainability issue, if we are going to create an environment in which the government can develop processes and thoughts to get us to the hard decisions that we are going to need to make, in a way that we develop the political consensus they need to make them – we need a new process. Evaluators, I think, are going to have a very important role in that process. And I am sort of giving some thought to how we engage together to try to figure out how best to move this country down the path of making the kinds of decisions it needs to make.

A couple of years ago, I was fortunate enough to be invited to a conference in Spain, sponsored by the Spanish Association of Parliamentary Lawyers, the theme of which was evaluating the impact of laws. And when I was invited, I thought, "They are talking about program evaluation". Uh-uh. They were talking about developing methodologies in such a way that when a program or law is *proposed*, evidence is brought to the table that it will work, before it is passed and before it is funded. Interesting thought – we do it a little bit in this country, in the regulatory arena, doing cost-benefit analyses and the like, but we never go back and look at whether or not those were correct.

So I think we need to open up the discussion beyond PART. I've been around a long time, folks – something PART-like may survive in the next administration, but it won't be what we have today. That's the nature of the political process, and at the heart of this is a political process. Michael mentioned earlier that he had "eyeballed" the programs in CDC to see if there was any relationship between funding decisions and the results of PART, and didn't see any. In the context of one of the engagements we did, we actually did an econometric analysis of that, and there is basically no correlation between the results of PART and funding decisions. But that's OK, because funding decisions, at their heart, are political decisions. And the idea of hiding behind a process like PART and trying to pretend it is objective, fact-based, etcetera, is a little misleading to what really needs to be done. And on that note, I think I'll conclude.

Questions and Answers

The closing section of this forum involved open questions and dialogue between members of the audience and panel participants. The following is a transcript of this Q&A session.

Q: Thank you all, this is very educational. I'm really interested in the fact that OMB examiners have so much power, and the fact that there is an opportunity to educate them through the interactions that you have as a Federal bureaucrat. But I'd also like to know if there is any recognition within OMB that it would be valuable for their examiners to have some understanding of evaluation if they are going to evaluate evaluation, as well as evaluate agencies.

Nancy Kingsbury: I would suggest that there has been some recognition of that need, in their willingness to accept the offer that the Federal Evaluators Network made to provide some fundamental information across examiners in different areas, to try to improve the consistency with which they come to the table to talk about evaluation. Again, the OMB budget process is intended to be a reflection of the President's priorities and the President's process, and so there is probably a relatively small extent to which you could influence it and move it much beyond where we've got it.

Q: I was intrigued by Nancy's comments towards the end, where you seemed to be saying that PART wouldn't survive like it is today, because of the politics. You also seem to call for a new process, and to think evaluators could play a role in that new process. Would you care to elaborate a little bit on what that might be? Are you suggesting at this point that evaluators might be thinking about the change in politics that will inevitably be coming down the road, in anticipating that? And do you have any thoughts about what directions we might want to be taking on that?

Nancy Kingsbury: When I had the idea a few hours ago to put that thought on the table, I had a feeling someone would be asking me what I meant by it, and that is a fair enough question. I keep coming back to what my boss is trying to do, and for those of you who haven't seen it, and haven't had an opportunity to do so, he has engaged in a partnership with a wide spectrum of outside organizations, starting with the Concord Coalition, Brookings, Heritage, across the political spectrum, in what we internally refer to as "the fiscal wake-up tour". And what he is trying to do is convince the American public that this problem of fiscal unsustainability is important and has to be resolved, but that it will take political consensus (with a small "p") to do that. He has got, I think, a good theme basically that until you get the public at large to understand what the problem is, the politicians themselves won't solve it.

So I think, in terms of what role could evaluators or anybody else play in terms of trying to struggle with, OK, suppose that we succeed and suppose there is a groundswell of the need to make some of these decisions. Some of them, like fixing the entitlement programs, are probably beyond even evaluators' skills, at least in the aggregate. But on a program by program basis, to look at programs about whether they are still needed now, and whether you could get constituencies and other parties like, say, local governments in the model that GPRA originally proposed, to begin to grapple with the need to make some of these hard decisions, it's just a thought that I decided to put on the table, and start thinking about myself, because I think if something is going to happen, this is the community that maybe has an edge on the kinds of thought processes and methodologies that might be useful in doing that. That's about as far as I've thought about it.

Q. I'm new to the field of evaluation, one year and counting, with a background in research, and I'm also new to AEA, and had a global question first, and then a more specific question. First, where would you like this to go? I mean, if this is a public issues forum on the PART, what do you hope to accomplish by this forum? And also, as an evaluator of one of the programs that I saw listed on one of the slides, a gear-up program in Ohio, I'm wondering what the implications of this might be for an evaluator of such a program – how might I look towards the PART as a tool for examining a program that may have already been evaluated by an OMB person, or a program that is yet to be evaluated?

Bill Trochim: Both of these are good questions. The first one, on what we hope to get from this forum, I'll take a crack at. As far as this forum is concerned, what led to this forum was an exchange that occurred on EvalTalk, shortly after ExpectMore.gov went live. I think it was readily apparent to anyone who was involved in EvalTalk at that point that ExpectMore.gov, which we showed a few slides of, triggered a variety of responses, some of them quite negative, about whether this PART effort and this website represented evaluation well, and whether as an evaluation organization, we ought to respond to it in some way, and whether we had a responsibility to take some position with respect to this kind of an endeavor.

At that point I was chair of the Public Affairs Committee, and we had a number of conversations where we felt we did need to respond to these kinds of situations. But it wasn't clear that we could go through the process of engaging the membership of AEA and try to formulate some policy with respect to PART, or that this was even the right time to do that. Instead, we thought that this would be a good time for us to learn about this, and to identify what seems to be going on here from an evaluator's perspective, and we thought that this kind of a forum would be a way for us to start this process. So this really was, and is, an experiment in mechanisms in learning for us – can we respond to situations that arise during the course of a year, identify ones that are particularly important for evaluators, focus on them in this kind of a forum, and then post the results of this so that other evaluators can also benefit from that.

Also, hopefully, I fully expect that some of the people at OMB will take a look at this forum, and review what we said here. I think Nancy already described an example of an interaction between the Federal Evaluators Network and OMB that had an impact on how they do some of the details in grading these programs and in terms of evaluation. And so I'm hoping that this will have an impact for the membership in terms of our learning about these things and engaging on these issues, and quite possibly we'll have some impact as a form of feedback to those policymakers who might read about this. So that's kind of the intent of this. And I guess the question that I heard you ask as well has to do with implications for evaluators who work with programs that have been through PART reviews, and I wonder if any of you would like to comment on that?

Michael Schooley: I think there are two things in the response. One, as a local evaluator and looking up at the list, and saying, "Hey, I'm doing local work and local evaluation with a program that might be an example of one of those competitive grant programs, a block grant or other grant type program. Looking at the PART evaluation, I think, would certainly be interesting and informative to you as a local evaluator, but is probably not going to tell you a whole lot about what you do on a day-to-day basis. It came up a little bit in a previous session on Federal assessments that PART reviews are done at a *program* level. What is happening locally, especially for many competitive grant programs or block grant programs, are *projects* that are implemented locally and individually. So looking at the evaluation of *programs* will be of interest to you, and it will have content for what you do in terms of the broad spectrum of your role, but it's not going to do a lot for the *projects* that you might be evaluating on a day-to-day basis.

One the other hand, looking at the PART tool I think is of use and value to local evaluators. I think the process and what they have pulled together is interesting and informative. I wouldn't say that

you would take it carte blanche and apply it to your local area, but there are certainly aspects of it that are of value. Some of the techniques might add to the pool of techniques you are using locally in the work that you are doing.

Nancy Kingsbury: I would agree with that. I think the fundamental question covers the kinds of questions that need to be thought about, and it might be enlightening just to look at those and say, OK, down at this project level, do we understand these things about, and do we have evidence of whatever kind, from whatever source, that could go towards answering those questions rather than trying to say yes or no to the fundamental substance of the question itself?

Q. I've enjoyed this panel very much, and have followed a lot of the Federal initiatives or performance measures, and I think my initial question has pretty much been answered in terms of what's the role of this Public Issues Forum to try to improve the PART process, which I happen to believe is likely to be a good thing if the right measures are chosen. But hearing what Nancy has said about how this is likely to not be continued in a new administration, and who knows what's going to happen after next week's election, but I'm thinking what might be the role of AEA and/or evaluators individually who know about a particular content area in trying to improve this process in the future, so that good performance measures are selected and evidence is weighed appropriately. It seems to me these are two key evaluation-oriented problems. Do you have any thoughts are to how we might contribute in the future?

Nancy Kingsbury: Let me say this. I think it will be interesting to see if anything comes out of the election that begins to mold where this is headed. A year ago at this conference, the debate around PART was almost entirely around the randomized control design that was out there, and Bill and I were talking before this session about whether that was going to be the topic of all the questions, and it hasn't been. That may be because of the urgency with which evaluators were reacting to the same thing that the GAO reacted to, which is that this is not an appropriate way to go around designing an evaluation tool for many, many programs, and have begun to see through the process itself in some kind of helpful way.

So going forward and looking ahead, I think if we're lucky, there will be a debate even in the campaign for President the next time around, that we may be able to feed into or help change. It's really hard to see that yet, but as soon as the election ends next week the presidential campaign will start. So I don't know the answer, really, but I do sense there is something of an opportunity.

Bill Trochim: I think that the little example that Nancy just gave us, of a group of federal evaluators bothering to pick up the phone and call OMB and arrange for a meeting and engage with them, actually had some pretty profound effects, given the cost of that effort. And as part of what I hope we take away from this is the idea that as members, as individuals, as evaluators, and perhaps even as an association, if we can find ways to begin to engage more directly in dialog around these things, we might be surprised to find the degree to which we might have some direct influence on the formation of legislation on the way in which these things are implemented, practiced and so on. At the very least, that ought to be part of what we are taking part in.

Q. I was really struck by a couple of comments. One was where Ted Kniker said couldn't we look at contribution instead of the causation, and I was also struck by Nancy Kingsbury's comments about trying to look at the interactions among Federal agencies for a wider strategic plan across Federal agencies. And I think that one of the fictions that underlies the PART is sort of like the old concept of union jurisdiction – that there is one union to one industry. But in practice it's not like that, and it isn't even necessarily clear it's a bad thing that it's not like that. So I think that there needs to be a lot more attention to whether or not the interaction and

interchanges between federal agencies are productive, useful, helpful, or whether it is simply redundancy. And the PART seems to basically say it's redundancy.

Ted Kniker: One of the things about the PART as it is currently implemented, I think, still promotes stovepiping of organizations, although in certain sectors there has been guidance from OMB that we need to broaden who is in particular programs. For example, with public diplomacy, the promotion of the U.S. image abroad, led by Karen Hughes, that the bureaus who worked on that were brought together, and there is talk about doing a united PART for that. But then there was also talk about, well, because there are other agencies that influence the image of the U.S. abroad, that they should be brought in as well, including the Department of Homeland Security, Department of Defense, Department of Justice, Peace Corps, and a number of these others.

So there is at least that recognition that programs and impact goes beyond just a particular agency. But one of the things that we have seen, and maybe an area where AEA can assist with, is in the Department of State, for example, the evaluation across the Department of State is not coordinated at all. PART was, but if my bureau was doing something on exchange programs, and how well particular programs worked, we didn't really involve any other programs that might contribute to that success. So I think that one of the things we need to look at is whether there are ways to begin to coordinate evaluation across agencies, if multiple agencies are involved with a particular program.

Bill Trochim: Well, I think our time is over, and with one final pun, thank you all for taking "PART" in this first AEA Public Information Forum!