

**Report from the Peer Review Panel
on activities within the
Planning, Analysis and Evaluation Unit
Office of Planning, Budget and Analysis
Office of Energy Efficiency and Renewable Energy
U.S. Department of Energy**

Team Members:

Kathryn Newcomer, The George Washington University (Chair)
Irwin Feller, The Pennsylvania State University (Retired)
Stephanie Shipman, Government Accountability Office
Kathleen Sedlak O'Brien, Environmental Protection Agency
Faith Lambert, EERE

Final Report
February 1, 2008

EXECUTIVE SUMMARY

The Office of Energy Efficiency and Renewable Energy (EERE) within the U.S. Department of Energy asked a panel of experts with relevant knowledge to review the evaluation activities of the Planning, Analysis and Evaluation Unit (PAE) within EERE's Office of Planning, Budget and Analysis. The panel consisted of two experts from academe, Kathryn Newcomer of the George Washington University and Irwin Feller of the Pennsylvania State University (Retired), and three experts from U. S federal agencies: Stephanie Shipman from the Government Accountability Office, Kathleen Sedlak O'Brien from the Environmental Protection Agency, and Faith Lambert from EERE. Short biographical statements about each of the team members are attached in Appendix A.

The panel reviewed materials sent to them and met from 8:30am to 6:30pm on December 11, 2007 in Washington, D.C. Staff from EERE briefed the panel, and then the panel deliberated about what they had read in advance of the meeting and heard that day. The panel members are in full agreement about their assessment of PAE activities. They formulated five major findings and five recommendations that appear in this report.

Major Findings:

1. The Peer Review Panel is very impressed with the progress that the Planning, Analysis and Evaluation Unit (PAE) has made since Jeff Dowd took direction of the PAE Evaluation Portfolio in expanding EERE's evaluation capacity and in the quality of the outputs produced thus far. We are also impressed by the Evaluation Portfolio leader's commitment to objective, technically rigorous evaluations.
2. PAE's mission requires that its staff perform measurement and monitoring assignments for EERE's Technology Development Offices on behalf of senior EERE management, and at the same time, that they develop a collaborative, collegial relationship to nurture and support the evaluation and performance measurement activities of these offices. PAE's emphasis on evaluation capacity building activities to date indicates that it is aware of the complexity of its role.

3. PAE is at a transition stage: its strategy of emphasizing capacity building, manuals and guides to best practice must now give way to the actual conduct of evaluations. This new strategy needs to be based on a specific prioritization of PAE's goals, objectives, and activities. Collaboration—getting buy-in—with the Technology Development Deputy Assistant Secretaries and Program Offices is needed for PAE staff to ensure that its activities help the programs obtain the evaluative information (or feedback) that the program managers need.
4. The PAE staff has an opportunity to take advantage of its placement at the corporate level within PBA to play a supportive, and even an advocacy role for obtaining the evaluative feedback that Program Offices need to help them meet their goals and objectives. To this end, PAE needs to help make the argument that analytical investments are needed to help program offices obtain needed information. While analysis and quantification evaluation of program cost and benefits and other impacts is important, PAE may gain support for evaluation from Program Offices by emphasizing both impact and process evaluations. Using evaluation efforts to improve programmatic learning and program improvement (formative evaluation) can be very beneficial to program managers, rather than having evaluation viewed exclusively as the means of reaching summative judgments about the worth or merit of programs. PAE has had success to date in the use of peer review, and the case can be made to program managers that evaluations focused on the implementation of programs, i.e., formative evaluations, such as efforts undertaken via peer reviews, can indeed promote programmatic learning.
5. The goal to develop a more consistent and integrated performance measurement framework is very laudable, and the performance measurement project has potentially high value in identifying and filling gaps in EERE's information system. The current effort could benefit from some redirection and focus; the current model is too comprehensive, and the data collection burden required may be overwhelming.

Recommendations:

1. To meet its goal of providing useful information to the programs, PAE should conduct its strategic planning in collaboration with program managers to identify which are the most strategically important evaluations to undertake.
2. PAE should continue to emphasize using evaluation to improve program performance; this focus should foster collaboration with program managers, and help in evaluation capacity building.
3. PAE's prioritization of evaluation studies should be based on recognition that the overwhelming portion of EERE's budget is allocated to applied research and development (R&D) programs, not technology deployment programs.
4. Evaluation of EERE's applied R&D programs should be conducted within the framework of OMB's R&D Investment Criteria. Relatedly, evaluation studies of EERE's programs should draw upon, and be benchmarked against empirical evaluations of comparable programs in other federal agencies.
5. We recommend that efforts be undertaken in PAE's range of project and performer selection mechanisms to (1) widen its network of potential performers; and (2) increase competition among them. For example, PAE might consider using RFPs or Task Orders.

INTRODUCTION

The Office of Energy Efficiency and Renewable Energy (EERE) within the U.S. Department of Energy asked a panel of experts with relevant knowledge to review the evaluation activities of the Planning, Analysis and Evaluation Unit (PAE) within EERE's Office of Planning, Budget and Analysis. The panel consisted of two experts from academe, Kathryn Newcomer of the George Washington University and Irwin Feller of the Pennsylvania State University (Retired), and three experts from U. S. federal agencies: Stephanie Shipman from the Government Accountability Office, Kathleen Sedlak O'Brien from the Environmental Protection Agency, and Faith Lambert from EERE.

The panel reviewed materials sent to them and met from 8:30am to 6:30pm on December 11, 2007 in Washington, D.C. Staff from EERE briefed the panel, and then the panel deliberated about what they had read in advance of the meeting and heard that day.

The panel is in full agreement about their assessment of PAE activities. The panel is extremely impressed with the level of commitment of the PAE staff, and with the quality of the work undertaken thus far. The entire panel is extremely supportive of the goals that the PAE staff has been pursuing thus far to build evaluation capacity within EERE, commends all efforts undertaken to date, and offers suggestions for how to achieve further growth now. The team formulated five major findings and five recommendations that appear in this report.

In this report first the major findings are listed, and below each finding additional comments offered by individual panel members are provided. Second, a summary of the quantitative assessments offered by panel members is presented, in accordance with the request from the EERE Office of Planning, Budget and Analysis's Planning, Analysis and Evaluation team. And third, recommendations are provided for consideration by the EERE Office of Planning, Budget and Analysis.

MAJOR FINDINGS:

1. The Peer Review Panel is very impressed with the progress that the Planning, Analysis and Evaluation Unit (PAE) has made since Jeff Dowd took direction of the PAE Evaluation Portfolio in expanding EERE's evaluation capacity and in the quality of the outputs produced thus far. We are also impressed by the Evaluation Portfolio leader's commitment to objective, technically rigorous evaluations.

- With modest personnel and financial resources, PAE has developed and published several high quality reports and manuals that provide guidance to the evaluation and performance measurement of EERE's operating units, provided technical support to these offices in initiating performance measurement and evaluation activities, and provided corporate-level support for EERE-wide performance information. Although there are plenty of references to "using evaluation to inform decisions," there aren't many examples of its use, or of evaluation efforts directly targeted at specific decisions.
- Jeff Dowd is to be commended for his commitment to seeking out the most relevant and up-to-date guidance on program evaluation and evaluation-capacity building. His passion for his work is evident, and impressive.

2. PAE's mission requires that its staff perform measurement and monitoring assignments for EERE's Technology Development Offices on behalf of senior EERE management, and at the same time, that they develop a collaborative, collegial relationship to nurture and support the evaluation and performance measurement activities of these offices. PAE's emphasis on evaluation capacity building activities to date indicates that it is aware of the complexity of its role.

- EERE is committed to a decentralized evaluation model. PAE is clearly intended to be a center for excellence in evaluation. Thus, PAE must work closely with program offices to encourage them to finance and participate in evaluation studies. The decentralized model implies that the individual program office should determine the purpose and scope of evaluations, while PAE provides

technical expertise to help the program office obtain credible, relevant information within available resources.

- EERE senior management leadership and commitment is a critical element in the successful accomplishment of the decentralized evaluation model. It is important for EERE senior leadership, with PBA/PAE playing a pivotal role, to develop an operational strategy for demonstrating their support for evaluation as an integral element of decision making in EERE.

3. PAE is at a transition stage: its strategy of emphasizing capacity building, manuals and guides to best practice must now give way to the actual conduct of evaluations. This new strategy needs to be based on a specific prioritization of PAE's goals, objectives, and activities. Collaboration—getting buy-in—with the Technology Development Deputy Assistant Secretaries and Program Offices is needed for PAE staff to ensure that activities help the programs obtain the evaluative information (or feedback) that the program managers need.

- PAE staff might need to reach out more to program managers to talk with them and help them to identify and prioritize their information needs (which could include information to respond to PART or similar systems, to formulate GPRA estimates for budget request formulation, to respond to inquiries from Congress and others about the full range of program outcomes and impacts, information on a program's operational effectiveness, and the like)
- PAE's goals are relevant to EERE's overall performance, but they appear to be too ambitious for its level of resources, especially if PAE is tasked with initiating and funding program evaluations. Indeed, most federal agencies do not have the resources to conduct evaluations of all their programs, so EERE needs a process for making decisions strategically with regard to evaluation; EERE, in addition to PAE, still needs an evaluative selection process to decide which programs they will evaluate, that is, to prioritize, to make wise funding choices.
- There's more discussion in the portfolio of different sources of information requests (e.g., Joule, PART) than of differences in types of information actually requested.

- At some points in the advance materials and briefings on December 11, program evaluation and performance measurement are discussed as if they are interchangeable. It might help to focus on which information is needed for monitoring on a routine basis (to answer annual/quarterly questions), versus which (needed) info is too expensive for regular measurement and should be the basis for a special evaluation study.
- PAE should consider how its goals and objectives fit in with those of EERE as a whole, and how evaluation activities can be structured to provide information needed to assess whether those goals have been met. The role and goals of PAE should be explicit and clearly understood by managers and staff throughout EERE.
- The development and use of program logic models and the extent to which the effort is being formalized in EERE is a big step forward. However, the logic models should emphasize the importance of **SHORT-TERM OUTCOMES FOR SPECIFIC INTENDED PROGRAM CLIENTS**, which provides a critical link between the R&D outputs and the long-term goals.

4. The PAE staff has an opportunity to take advantage of its placement at the corporate level within PBA to play a supportive, and even an advocacy role for obtaining the evaluative feedback that Program Offices need to help them meet their goals and objectives. To this end, PAE needs to help make the argument that analytical investments are needed to help program offices obtain needed information. While analysis and quantification evaluation of program cost and benefits and other impacts is important, PAE may gain support for evaluation from Program Offices by emphasizing both impact and process evaluations. Using evaluation efforts to improve programmatic learning and program improvement (formative evaluation) can be very beneficial to program managers, rather than having evaluation viewed exclusively as the means of reaching summative judgments about the worth or merit of programs. PAE has had success to date in the use of peer review, and the case can be made to program managers that evaluations focused on the implementation of programs, i.e., formative evaluations, such as efforts undertaken via peer reviews, can indeed promote programmatic learning.

- PAE staff may want to educate program managers on the various types of formative evaluation strategies that they may employ to help them understand how well programs are performing, that is, evaluations of implementation, process and quality of outputs, as well as outcome and impact evaluation options.
- The majority of EERE's research support is directed at applied R&D projects. A considerable body of theoretical and empirical research about the effectiveness and impacts of such programs, both those sponsored by DOE and other Federal agencies, e.g, NIST, has developed in recent years. Relatedly, the research community capable of conducting expert work on such programs has grown considerably. Given these developments, and consistent with OMB's emphasis in its R&D Investment Criteria memos on the use of competitive, merit based review processes to select projects and performers, PAE should be able to find talented contractors (see Recommendation #5).

I'd like to see more on the types of decisions made at different organizational levels (EERE, program, project) matched to the types of info those actors are looking for, e.g., barriers to project success, balance and relevance of portfolios, industry/market support for innovations

- I'd like to see more on the types of decisions made at different organizational levels (EERE, program, project) matched to the types of info those actors are looking for, e.g., barriers to project success, balance and relevance of portfolios, industry/market support for innovations.

5. The goal to develop a more consistent and integrated performance measurement framework is very laudable, and the performance measurement project has potentially high value in identifying and filling gaps in the EERE's information system. The current effort could benefit from some redirection; the current model is too comprehensive, and the data collection burden required may be overwhelming.

- The experts who have constructed the performance measurement framework are to be commended for their extensive review of the

literature and practice. They have been careful to identify well respected measures, and their logic is strong and defensible.

- While we appreciate that the comprehensive list is meant to offer a menu, the problem is that it will be received by an audience in which there is some “metric fatigue”, and seeing so many measures can be intimidating.
- We have generally found that frameworks that aim to identify comprehensively all the possible measures of program performance and influential factors tend to scare off both potential participants and audience for an evaluation. I’d limit the framework to naming the boxes of major components of performance rather than listing potential measures within each box. Then, I’d provide 1-2 illustrative examples of designs that would use a couple measures from a couple boxes to answer one or two commonly asked evaluation questions. Refraining from providing a long list of measures would allow the program offices to choose which measures to include in their evaluation studies while at the same time ensuring that evaluation quality and rigor remain high.
- How important is it that EERE programs use common measures? Will their efficiency, quality, or effectiveness be directly compared? If infrequent, would such program comparisons be more appropriately tailored on a case-by-case basis, as the need arises? Allowing program offices greater latitude in customizing the design of their individual evaluation studies would ensure that their specific information needs are met, and encourage continuing collaboration with PAE.
- Note: it is very difficult to judge the management and quality of the Performance Measurement project because there’s not much detail about how it would actually be carried out, and thus about feasibility, cost, reliability, validity, suitability for trend analysis, etc.

QUANTITATIVE SCORING

The Peer Review Panel chose to aggregate their scores on the various facets they were asked to rate. The scores appear in Table 1. In submitting its scores, the Panel notes that it does not consider the numerical scoring format an especially useful or helpful way to provide feedback to the PAE staff.

Given the Panel's agreement about the high quality of PAE's leadership and its accomplishments to date, some of the lower numerical scores simply reflect individual panel member reservations about specific actions or plans or a lack of information on which to evaluate, and should not be interpreted as providing different or more valuable feedback than the comments made earlier in this report.

Please note that the scores given on the four sets of performance measures are given on the sets of measures the panel was asked to review, which are program management, technological readiness, market readiness and ultimate outcomes, rather than the four categories provided on the scoring sheet.

Please take note of the following contextual comments when interpreting the numbers provided in Table 1.

First, the Panel recognizes the complex situation in which the PAE staff members find themselves in terms of the variety in clients and consumers for program evaluation and performance measurement. We do not have any magic bullets to offer, but we recognize that the tasks facing the PAE staff demand an incredibly high level of collaboration – and this is really time-intensive.

Second, if we were asked to rate the same items for practically any other federal agency, our numbers would be the same or lower! Some other agencies may be farther ahead in performance measurement, but some of these are suffering from metric fatigue. Our scores for outcome measures mostly reflect our fear that the measures would be expensive and challenging to capture, and burdensome for the programs offices.

And third, unfortunately there are not any huge success stories in other federal agencies that we can recommend in terms of the similarity of their

challenges and the transferability of their successes. Measuring such elusive concepts as market readiness will not be easy anywhere.

We take the term performance management to mean that program managers and top management actually use the performance measures to inform decisions and help guide strategic planning, and that is *simply not commonplace* across the federal government. PAE can provide guidance and support to capture great measures, but it will be virtually impossible for the PAE staff to ensure that the measures are truly used by managers.

We are cautiously optimistic about evaluation capacity building and performance measurement that informs management, and we recognize that the PAE staff is anxious to learn from experience in other agencies. There are some good examples: The Center for Disease Control (CDC) within the Department of Health and Human Services and the U.S. Coast Guard both have enviable records in evaluation and performance measurement, but their goals are easier to operationalize than those of DOE.

More in line with EERE's mission, the National Institute of Standards and Technology's (NIST's) Advanced Technology Program (ATP) has had a major program evaluation component. An overview, including citations from most studies, is contained in the report that Rosalie Ruegg and Irwin Feller did for ATP, "A Toolkit for Evaluating Public R&D Investment: Models, Methods, and Findings from ATP's First Decade," (GCR03-857) That report received the American Evaluation Association's award for best professional study.

The Manufacturing Extension Partnership Program (MEP) is another NIST program that may provide some relevant guidance. MEP has/had a major evaluation component, including an external advisory board. You may wish to check the web for Phil Shapira's writings about the MEP. In addition, the National Science Foundation (NSF) has had a long standing, ongoing evaluation of its Industry-University Cooperative Research Program. You may wish to check the web for work by Denis Gray of North Carolina State University about that program. NSF also had a task contract with SRI International to evaluate its Engineering Research Program. Most of SRI's studies ended up as NSF in-house reports, but some ended up in peer reviewed journal articles, e.g., an article on Research Policy by Irwin Feller.

Table 1: Quantitative Scores Given to the Performance Measures Project and the PAE Evaluation Portfolio

Category	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5	Mean
Performance Measures: Relevance	8.5	6.5	7.5	8	7	7.5
Performance Measures: Management	6	3	5	4	3.5	4.3
Program Management Measures	7	5	6	7	5	6
Technological Readiness Measures	7	4	6	7	5	5.8
Market Readiness Measures	6	4	5	7	4	5.2
Ultimate Outcome Measures	3	1	3	3	4	2.8
Overall Rating on Performance Measures	6.5	4	5.5	6	5	5.4
Portfolio Relevance	8	5	8	8	6	7
Portfolio Management	7.5	3	6.5	6	6.5	5.9
Portfolio Quality	7	5	7	7	7	6.6
Overall Portfolio Rating	7	5	7	7	6.5	6.5

RECOMMENDATIONS:

1. To meet its goal of providing useful information to the programs, PAE should conduct its strategic planning in collaboration with program managers to identify which are the most strategically important evaluations to undertake.
2. Continue to emphasize using evaluation to improve program performance; this focus should foster collaboration with program managers, and help in evaluation capacity building.
3. PAE's prioritization of evaluation studies should be based on recognition that the overwhelming portion of EERE's budget is allocated to applied R&D programs, not technology deployment programs.
4. Evaluation of EERE's applied R&D programs should be conducted within the framework of OMB's R&D Investment Criteria. Relatedly, evaluation studies of EERE's programs should draw upon, and be benchmarked against empirical evaluations of comparable programs in other federal agencies.
5. We recommend that efforts be undertaken in PAE's range of project and performer selection mechanisms to (1) widen its network of potential performers; and (2) increase competition among them. For example, PAE might consider using RFPs or Task Orders.

Appendix A

Kathryn Newcomer (Team Chair) is a professor at the Trachtenberg School of Public Policy and Public Administration at the George Washington University where she is also the Director of the Midge Smith Center for Evaluation Effectiveness, home of the Evaluators' Institute, and she is the Director of the PhD in Public Policy and Administration program, and Associate Director of the School. She teaches public and nonprofit, program evaluation, research design, and applied statistics. She routinely conducts research and training for federal and local government agencies and nonprofit organizations on performance measurement and program evaluation, and has designed and conducted evaluations for several U.S. federal agencies and dozens of nonprofit organizations.

Dr. Newcomer has published four books, Improving Government Performance (1989), The Handbook of Practical Program Evaluation (1994, 2nd edition 2004), Meeting the Challenges of Performance-Oriented Government (2002) and Getting Results: A Guide for Federal Leaders and Managers (2005), a volume of New Directions for Public Program Evaluation, Using Performance Measurement to Improve Public and Nonprofit Programs (1997), and numerous articles in journals including the Public Administration Review. She was identified as one of the top 25 evaluation experts in the country in 2001 by the American Journal of Evaluation. She is a Fellow of the National Academy of Public Administration, and currently serves on the Comptroller General's Educators' Advisory Panel. She served as President of the National Association of Schools of Public Affairs and Administration (NASPAA) for 2006-2007. She has received two Fulbright awards, one for Taiwan (1993) and one for Egypt (2001-04). She has lectured on performance measurement and public program evaluation in Ukraine, Brazil, Egypt, Taiwan, and the UK.

Dr. Newcomer earned a B.S. in education and an M.A. in Political Science from the University of Kansas, and her Ph.D. in political science from the University of Iowa.

Irwin Feller, The Pennsylvania State University (Retired) is a senior visiting scientist at the American Association for the Advancement of Science. He is also emeritus professor of economics at the Pennsylvania State University, where he served on the faculty for 39 years, including 24

years as director of the Institute for Policy Research and Evaluation. His current research interests include the economics of science, the evaluation of federal and state technology programs, and the university's role in technology-based economic development. He has been a consultant to the President's Office of Science and Technology Policy, as well as numerous other government, not-for-profit, and private sector organizations. He has served as chair of the National Science Foundation's Advisory Committee to the Assistant Director for Social, Behavioral and Economic Sciences, and as both chair and member of several National Research Council Committees. He has a B.B.A. in economics from the City University of New York and a Ph.D. in economics from the University of Minnesota.

Stephanie Shipman, U.S. Government Accountability Office is an Assistant Director of the Center for Evaluation Methods and Issues in the Office of Applied Research and Methods. At GAO, she has evaluated various programs serving children and families, including welfare-to-work initiatives, and researched approaches to meeting congressional information needs. Over the past several years she has directed studies of federal agencies' performance measurement and program evaluation activities, and methods for solving various analytic challenges in program performance assessment. She has consulted with numerous federal agencies and foreign government auditing agencies on program evaluation policies and practice, and currently serves on the American Evaluation Association's Evaluation Policy Task Force. Dr. Shipman is a founding member of and coordinator for the Federal Evaluators group, an informal network of evaluation officials.

Dr. Shipman received her A.B. degree from Princeton University and a Ph.D. in Educational Psychology, Measurement and Evaluation, from Teachers College, Columbia University.

Kathleen Sedlak O'Brien, Environmental Protection Agency, has worked at the U.S. Environmental Protection Agency since 1989. She is currently Director of the Office of Planning, Analysis, and Accountability in the Office of the Chief Financial Officer. Kathy and her colleagues are responsible for designing strategies and managing Agency-wide implementation efforts to strengthen performance measurement and management under the Government Performance and Results Act (GPRA), other key management legislation and implementing guidance.

Responsibilities include managing EPA's Program Assessment Rating Tool reviews, development of EPA's Strategic Plan, consideration of performance results in formulation of EPA's Annual Plan and Budget, and development of EPA's annual Performance and Accountability Reports. Among other activities, Kathy's office works with EPA program and regional offices to help them develop more meaningful strategic and annual performance measures for use in managing programs and informing Agency decisions. Kathy's office also engages with Agency state and tribal partners to better align and strengthen joint EPA/state/tribal planning and priority setting.

Before joining EPA, Kathy worked for six years at the U.S. Office of Management and Budget on implementing government-wide management reforms. Kathy has a Master of Public Administration degree from the George Washington University and a Bachelor of Arts degree in Spanish from the University of Massachusetts/Amherst.

Faith Lambert, State Energy Program, Office of Weatherization and Intergovernmental Programs, U.S. Department of Energy, has worked in energy policy and programs since 1973, soon after the Arab Oil Embargo, when she joined the Office of Energy Conservation at the Department of Interior. She began in the Public Affairs Division, and then moved into the State and Local Programs area in the Federal Energy Administration, where she participated in the early planning and design of the Schools and Hospitals Grant Program. Over the years she has been active in policy development and program operations in the Department of Energy's programs for States, local governments, and institutions. Ms. Lambert has been a part of the DOE State Energy Program Team for the past 20 years. For the past 15 years she has also been active in the evaluation of energy efficiency and renewable energy programs, serving on the Planning Committee and the Board of Directors for the International Energy Program Evaluation Conference. In addition, since 1986, she has been team lead for DOE's monitoring and oversight of the oil overcharge funds received by the States as a result of cases stemming from the Arab Oil Embargo.