CSREES' RESPONSE TO GPRA

Donald E. Nelson

Abstract: As its name implies, the USDA’s Cooperative State Research, Education, and Extension Service (CSREES) conducts its research, education, and extension programs through state cooperators—principally through the land grant colleges. As a federal agency, CSREES' programs are subject to the same Government Performance and Results Act (GPRA) requirements as all other federal agencies. GPRA sets forth, in a very business-like fashion, a planning, implementing, reporting, and subsequent funding cycle, through which future agency funding will be based upon agency performance as measured by specific performance indicators. All agencies which fund basic research are struggling to identify indicators which accurately measure performance. CSREES' organizational structure adds some additional difficulties not faced by most other agencies. CSREES provides only a portion of the funds (sometimes a very small portion) for the programs it supports. Should whatever indicators that are measured be reduced in proportion to the proportion of federal funding? CSREES is not a direct-line agency, so the extent of its authority to require its cooperators to measure performance may be questioned by some. Is it even possible to develop a nationwide reporting system that makes sense? Who should pay for the cost of measuring performance? Where the federal share is relatively small, it may be insufficient to cover the cost. As the above implies, CSREES is struggling with GPRA. Progress has been made in the area of general strategic planning for GPRA, but the precise details of what to measure and how have yet to be agreed upon. It does not help that perceptions of what is expected are not clear.

Introduction

If you are involved in any way with research, teaching, or extension programs that receive some federal funding, you may be affected by the Government Performance and Results Act (GPRA). If you are a cooperator of USDA’s Cooperative State Research, Education, and Extension Service (CSREES), you may be asked to measure your performance using GPRA performance indicators as a part of CSREES' overall GPRA plan. If you and CSREES do a good job of measuring your performance and if your performance results in benefits to clientele, you will have contributed favorably toward building a GPRA data base that will be available for use by Congress and others who make decisions relative to what gets funded and what does not.

Past Efforts to Evaluate Programs

Program evaluation is not a new concept to government. What appears to be somewhat new, however, is the apparently serious intent by decisionmakers to actually use the results of evaluations of past programs to make funding decisions for future programs. Although space does not permit a lengthy discussion of all past schemes for evaluating federal programs, the remarks of Extension Service officials Ed Kirby and George Hyatt (1974) before the OECD in Paris, in which they noted little success in developing output measures, provide a brief historical perspective:

---

In the United States, Extension like other public agencies, has been under increasingly greater pressure to develop means of measuring or evaluating progress in order to more effectively account for use of public funds.

This was reflected in the establishment of the Programming, Planning and Budgeting System (PPBS) employed at the national level in the 1960’s and the establishment of planning and evaluation staff offices and reporting systems in the 1970’s, designed to implement Management by Objectives (MBO) ...

In Extension we have had little success in developing measures of output within this framework, as has been the case in other agencies not involved in the production of products or services ...

To date, the most effective means of evaluating programs have been the observations and conclusions of private citizens about the benefits they receive from Extension professionals and the justifications largely involve generalized statements about program benefits and specific examples of benefits which individuals or groups have received from some part of the Extension program ...

As yet, we have not developed an effective system of measuring accomplishments within the Extension Management Information System ... The continuing pressure on budgets and competition for use of public funds will provide additional incentives for Extension Administrators and Managers to develop more effective means of evaluating and reporting on the use of public funds. Important steps forward will likely come only with large expenditures of funds on research to develop the necessary methodology to effectively measure the impact of the program.

Note, in particular, the last sentence relative to the likely high cost of developing the necessary methodology. Twenty years later, some progress has been made, but much remains to be done.

The Kirby and Hyatt quote is made within the context of extension programs, where at least there is a specific clientele group to whom extension programs are visible. How visible to the general public are the results of research and education programs? What would be the “observations and conclusions of private citizens” relative to research and education programs, if that was to be the measure?

Before leaving the subject of past attempts at program evaluation, it would be well to repeat Steele’s (1975) useful distinction between “project” and “program” evaluation: “The first is how well you do what you do. The second is how well you chose what you should do.”

Drawing upon Steele’s insight, I (Nelson 1983) commented about the difficulty in selecting what you do in order to not incur opportunity costs:

A “project” may achieve the results intended. Steele argues this does not necessarily mean the “program” (of which the “project” is a part) is a success. Even if all projects undertaken reached their expectations, other projects might have made a greater contribution. This is the opportunity cost of not having used those same resources in some other project which might have resulted in a more successful program. One cannot dispute the logic of this approach, but it is of no help in selecting the “right” project unless one possesses the ability to pre-evaluate all potential “projects.” (Post-evaluation is difficult enough).
Even if the project is well selected, and serious attempts are made to evaluate it, Forest (1976) quotes evaluator Douglas Sjogren's claim that evaluations are mostly ignored:

The situation is almost a comedy. An "evaluator" is hired to play the role while both he and the project director know the evaluation activity will make little difference to anybody ... While the evaluation role is being played, the real evaluations that do make a difference are being made by the project staff and the constituents. The evaluator is often not even aware of their judgments.

A somewhat opposite view is expressed by Mushkin (1973), who thinks evaluation has been used but is being misused by decisionmakers:

Evaluation is being used as a decisionmaking tool more than it warrants, especially in light of the present state of scientific understanding and of the expectations placed upon evaluation by the public and its officials.

The major evaluations of social programs have yielded negative or "no clear" findings, and these findings have come to be interpreted as saying "The results anticipated from social policies are not being achieved." Health is not better, nor is housing for the poor ... etc.

In multiple regression analysis in which the dependent variable that is being explained is a fairly general factor (such as crime rate, death rate, etc.) each of the many independent variables that may be included in an analytic model is likely to have small power to explain variation. Further, analysis of variance of regression does not methodically produce the causal connections required to examine any hypothesis made regarding the programs. Yet the results of regression analysis are being applied as if they in fact examined those causal relationships and concluded in the negative.

Whether ignored or misused, it is rather clear that evaluations of public programs in the past have not provided the kinds of information that decisionmakers feel they need to make funding decisions.

GPRA to the Rescue

The Government Performance and Results Act of 1993 (GPRA) is the most recent attempt to provide a better framework for government operations. On August 3, 1993, President Clinton said: "The Law ... requires that we chart a course for every endeavor ... see how well we are progressing, tell the public how we are doing, stop the things that don't work, and never stop improving."

More specific purposes of GPRA were given as:

- Improve the confidence of the American people;
- Initiate program performance reform;
- Improve program effectiveness and accountability;
- Help managers improve service delivery;
- Improve congressional decisionmaking;
- Improve the internal management of the Federal government.
CSREES' Response

Shortly after GPRA was enacted, a reorganization of the U.S. Department of Agriculture combined the former Cooperative State Research Services (CSRS) with the former Extension Service (ES) into a new agency—CSREES. Previous planning to implement GPRA within the separate agencies was scrapped in an attempt to go forward with a unified GPRA strategy for CSREES. Although now a combined agency at the federal level, the individual functions of research, education, and extension were still viable in the cooperating universities.

In the spirit of unity, however, an attempt was made (in a planning meeting in Memphis in December, 1995) to identify goals for the new agency which would be all inclusive of the individual functions, and to identify a small number of general performance measures that could be applied to all three functions. Although the planners were able to identify five goals and a small number of subgoals under each, they were not able to identify a small number of performance measures that could be used simultaneously to measure performance of research, education, and extension. Each individual planner at the meeting who had a vested interest in a specific program (water quality, for example) wanted the group to identify a performance measure that spoke directly to that vested interest. But since the group had to select only a few performance measures, the measures selected had to be broad (“increased scientific knowledge base,” for example) to include more than just a small segment of the total program. But that made the indicator too broad to be able to relate it to an individual budget line item in the agency’s budget. There was considerable uncertainty at the time of the Memphis meeting as to how closely performance measures would need to relate to budget line items. An argument was made that performance measures only needed to be able to show some “causal relationship” to budget line items, but, given the intent of GPRA, it was difficult to see how general measures would be satisfactory and useful to decisionmakers faced with allocating funds between specific line items.

This uncertainty was apparently cleared up when USDA’s Chief Financial Officer issued GPRA guidance on February 15, 1996:

All Federal departments are to prepare an Annual Performance Plan covering each program by activity line set forth for that agency in the President’s Budget ...

Performance Plans should link annual performance goals and measures ...

Performance goals are the target level of performance (output or outcome) expressed as a tangible, measurable result ...

Performance indicators are specific values ... used to evaluate the output or outcome of a performance goal.

GPRA planning thus far in CSREES may have dealt too much with generalities, whereas the actual requirement may be much more specific than was assumed at the beginning. If the more specific is really required, it is yet to be seen how CSREES will deal with establishing the massive, nation-wide data collection system that would be needed if specific performance indicators are required for each of the 52 line items in the CSREES budget. (Some of these line items are for formula funded or grant programs, each of which supports very diverse activities within the single line item, so many indicators would be needed just for McIntire-Stennis or Smith-Lever funding.)
Opportunity Costs

"Proving" that CSREES programs benefit participants (or hurt them) requires random selection and control groups, just like any good research. But how can one select at random without reducing the cost-effectiveness of the program below what it would have been if willing and anxious participants had been selected? Or what willing and anxious participants will give up program benefits in order to serve as a control? How can we permit careful (non-random) selection of audiences, make the program available to all (no controls), and still evaluate with credibility? Difficult as post-evaluation of CSREES programs is, we really should be able to pre-evaluate so that we can select those programs that will provide the greatest benefits.

Each individual CSREES program results in an opportunity cost unless it is perfect—the very best that could have been done. Each CSREES cooperator, therefore, accumulates additional opportunity costs with each successive "less than perfect" event. Good evaluation can help to reduce the grand total of opportunity costs by enabling each worker in the system to conduct more "perfect" and "near-perfect" educational events. Evaluation leads to efficiency. Evaluation is a tool which helps "perfect" program selection and delivery. If we never measure the results of our programs, we can not possibly know how they could have been improved or if they were properly selected in the first place. It is just possible that we might have accomplished much more if we had put our scarce resources to an entirely different use or used them in an entirely different way.

Opportunity costs can, like a two-edged sword, cut in both directions. Now that you accept that a poorly selected or poorly conducted program has an opportunity cost, there is another opportunity cost that you should also consider. Consider the opportunity cost associated with the resources taken away from CSREES programs in order to conduct evaluation studies. Some of these same resources could have instead been used to conduct new programs rather than to evaluate past programs.

If the results of an evaluation produce less benefits than would have been produced by some additional CSREES program, then the opportunity cost of having done the evaluation instead of the additional CSREES program is the loss of the additional benefits (over and above the benefits from the evaluation) which could have been gained had the resources used in the evaluation been used instead to conduct the new program.

Further, if you "waste" resources by evaluating an already "perfect" program, you don't learn anything that will help improve that specific program, since you can't improve on perfection. So if you are so fortunate as to always conduct only "perfect" programs, then you never incur any opportunity costs. If so, you have blown your opportunity to be perfect, anyway, by choosing to evaluate your "perfect" programs, because you could have spent those same resources conducting additional "perfect" programs rather than "wasting" them on "unproductive" evaluations which only give you a pat on the back for your excellent work, and perhaps, good data for GPRA. Here we have a real dilemma: The better our programs are, the lower the opportunity costs of the programs themselves (good), but the greater the opportunity costs of evaluating them (bad), since the evaluation will uncover little potential for improving our already-good programs. Conversely, the poorer our programs are, the greater the opportunity costs of the programs themselves (bad), but the lower the opportunity costs of evaluating them (good), since the evaluation will suggest numerous ways that future programs can be improved. Damned if we do and damned if we don't!

Will GPRA require us to spend so much money to prove we are doing the right thing, and so much money to establish that we have, indeed, performed, that we will have wasted many of the resources available to us that would otherwise have been available to produce more benefits for the American
public? Does it benefit the American public by spending tax dollars to measure performance at the expense of programs to benefit the public?

Some Questions of Who, What, Why, When, and Where

Who Should Evaluate?
Involved the issues of credibility, motivation, and expertise. Can the CSREES specialist who conducted the educational programs objectively evaluate the same? He/she may be strongly motivated to do so or not motivated at all. He may be too engrossed in the program to be able to look at it objectively “from the outside.” On the other hand, his immediate supervisor may be even less objective, only wanting to show great benefits in order to better compete for funds within the agency. An outside evaluator lends credibility if given a free hand, but it is difficult for him/her to not be influenced by the CSREES specialist and others involved in the program. Influencing the outside evaluator may become the most important educational program of the CSREES specialist.

What Should be Evaluated?
Involved issues of selectivity. Do we evaluate all programs? No, that would be too expensive, but will small programs that are not evaluated fall by the wayside because there is insufficient GPRA data to support the line item? Which ones, then, and who selects them? It comes as no surprise that an objective evaluation of randomly-selected programs turns out differently than an objective evaluation of purposefully-selected programs. Given a choice, most agency managers select their best programs to be evaluated so the evaluations will show them to be effective managers. The lack of objectivity in selecting programs to be evaluated is probably a much more serious bias than the lack of objectivity in evaluating programs. The best possible situation for a program manager is to be able to select his/her best programs and evaluate them internally and non-objectively. The manager is thus sure to emerge “smelling like a rose.” Most evaluation begins by selecting programs already believed to be effective, rather than objectively selecting programs at random, or purposely selecting programs of questionable value.

Why Evaluate Programs?
Issues here involve custom, curiosity, program improvement, ego, legal requirements, proof of benefits, return on investment, salesmanship for future programs, and feedback of research needs. Some CSREES specialists routinely evaluate their programs because it is customary to do so. Such evaluations often involve asking the audience immediately at the close of the program to indicate how well the program met their needs, if the facilities were comfortable, etc. This kind of evaluation is accepted in some organizations as routine, but it is most applicable to extension programs where a specific audience is involved. A training organization may even have a standardized instrument which they use to evaluate all the programs they conduct.

When to Evaluate?
A program raises the issue of the time value of money, and how far down the chain of subsequent events one should look to claim benefits. If the benefits from a program are spread out over time, those which occur first are worth more (in present-value terms) than those which occur later. To be completely accurate, the evaluation should measure benefits periodically until there are no more benefits which can be attributed to the program. Then some suitable interest rate must be used to convert all the benefits to a common point in time, so they can be compared to the total costs. When benefits are in dollars, this transformation can be easily accomplished mathematically, but what, when benefits are in some other units? Can one legitimately discount the value of a future recreational day, for example, using the same percentage value that one would use to discount a future dollar?
Where to Evaluate?
A program is not as evident as it may seem at first blush. Audiences for educational programs carry
the concepts away with them, and sometimes re-disperse these technology concepts through programs
of their own. Some publications which result from educational programs receive statewide,
nationally, or even worldwide distribution. Sampling the benefits only in those areas believed to
be of most interest in the program may miss or underestimate benefits which happened
to occur elsewhere. One obvious solution to where is to measure benefits at the point in time at which
the audience is still all together, but (as noted above) that may be too soon to measure actual changes
that have taken place as a result of an educational program and it works best for group educational
programs. A short-cut is to ask if and when those who benefit anticipate using the technology they
learned, and then follow-up by mail or telephone to see if and when they actually used the newly
learned technology.

Some Thorny Problems

- End use of GPRA—Will congress use GPRA instead of more traditional fund allocation
  processes? That is the $64,000 question.
- Accurate measurement of environmental conditions—Will environmental monitoring series
  be established that will be sufficiently accurate to detect the benefits of individual agency
  programs, line item by line item?
- Cost—How to pay for GPRA? What is a cost-effective level of evaluation? What level of
  precision is needed?
- Research—Can one use annual performance indicators to measure intermediate results of a
  long-term project, or a project whose benefits are far into the future?
- Employee performance—Is it proper to use the same performance measures to evaluate
  employee performance for merit raises?
- Skilled evaluators—Can staff be trained? Should in-house staff be used, or are they biased?
  Are outside staff available? Can uniformity be maintained for a credible, nationwide system?

Conclusions
This is definitely an interim report on GPRA in CSREES. It raises more questions than it answers.
Reasonable progress has been made toward a CSREES strategic plan, but little progress has been made
in the specifics of implementing a nationwide system through which CSREES cooperating
institutions contribute toward developing a CSREES GPRA database. Performance measures for the
outcomes of basic research are especially difficult to identify and measure. It almost requires a second
and more expensive research project to measure what happened as a direct result of the first, and
perhaps a third to measure if the second correctly measured the first.

Literature Cited
advisory) work in the United States." Page 5 in Proceedings of the Meeting of Directors of
Agricultural Advisory Services, OECD. US Department of Agriculture, Extension Service.

13-22.
