Asking Science to Measure Up

While scientists worry about the next budget, a new law that requires agencies to measure their performance could fundamentally change how research budgets are put together

Measurement is an essential element of science. Now, for the first time, the government is taking the idea of measuring performance and applying it to science itself. A little-publicized law, the Government Performance and Results Act (GPRA) of 1993, requires each federal agency to develop a strategic plan describing its goals, devise yardsticks to measure its progress, and tie that performance to its annual budget request. The law covers every type of science—from the most basic research supported by the National Institutes of Health (NIH) and the National Science Foundation (NSF) to the applied research conducted by the Department of Defense laboratories and the National Oceanic and Atmospheric Administration (NOAA).

This demand is posing quite a challenge for most research agencies. "We know that research is not amenable to the type of measures that work for Social Security," said Michael English, a senior management analyst for the Department of Health and Human Services (HHS), which includes NIH, to a meeting of 42 government research managers from more than a dozen agencies who crammed into a small room last month to discuss the new law. "Our budget people are throwing up their hands; nobody knows what to do," he said. "But the fact of the matter is that we have to do it."

Although the law doesn't take full effect for a couple of years, nobody is expecting the new Republican-controlled Congress to come to the rescue. One reason is that the idea of holding agencies accountable to taxpayers is a central tenet of the new majority party. Another is that GPRA was championed by Senator William Roth (R–DE), the new chair of the Senate Governmental Affairs Committee, which will oversee implementation of the law. "We're not asking an agency to promise that it will cure cancer in 3 years," says John Mercer, the committee's new majority counsel. "But they are going to have to do something more than say, 'Just give us the money.' GPRA points the government down a certain road that requires strategic planning, although how far down that road we can go remains to be seen."

Blazing a path

The challenge facing research managers is to take the first steps down that road by developing goals and performance measures for the projects they want to fund. "We're looking for a set of assessment options that are widely understood by the research community and have credibility with the public," says M. R. C. Greenwood, associate director for science at the White House Office of Science and Technology Policy (OSTP). Even if such measures can be developed, however, some government officials worry that the exercise could backfire if Congress or the Office of Management and Budget (OMB) decides to ax the budgets of agencies that fail to meet their goals. Indeed, Representative John Kasich (R–OH), a leader in past Republican efforts to reduce federal spending, said last year that Congress should play a larger role in overseeing GPRA with an eye toward "reallocating" agency budgets to bring them in line with the results of performance reviews.

NOAA Quantifies Its Mission

National Oceanic and Atmospheric Administration

Program/Indicator	FY 94	FY 95
1. Long-term climate change research		
New balloon-launching stations to measure tropospheric ozone trends	3	6
% commercially viable CFC substitutes evaluated for global warming potential	0%	25%
% completion of initial assessment of aircraft emissions on global warming	0%	10%
2. Short-term forecasts and warnings		
Severe thunderstorm warnings		
Lead time (minutes)	14	15
Accuracy (percent)	72%	75%
Temperature		
Correct forecast (percent)	83%	84%
Accuracy of forecasting onset of freezing temperatures (percent)	68%	72%
3. One-year lead time climate forecasts		
Dynamic forecast model operational	25%	50%
U.S. temperature and precipitation % correct forecasts	15%	15%*
Lead time (years)	0.25	0.50

* No change due to reduced funding in 1995

SOURCE: Excerpt from the agency's FY 95 performance plan submitted to the Office of Management and Budget

In the face of such concerns, Greenwood is urging research agencies to develop their own performance measures now, before requirements are imposed on them. "It's important that we're ahead of the curve on this, not responding to other people's ideas," she told a workshop last November. Some agencies, including NSF and NOAA, have already heeded the call (see table). Those agencies are conducting pilot projects to road-test their ideas on performance measures before they go into effect. NSF had a running start because it had already developed strategic plans for a congressional committee, but officials say the pilots also give them a chance to explain NSF to a larger audience. "We wanted to grapple with the pluses and minuses as soon as possible, as well as see what we could learn without overburdening the community with demands for data," says Judith Sunley, assistant to the director of NSF for planning. "It will also give us a chance to tell our story to the public, by making clear what NSF is doing with their money."

But other agencies—in particular NIH—are moving much more cautiously. Officials at the National Cancer Institute (NCI) are mindful of the negative fallout from President Nixon's war on cancer in the early 1970s—an exercise in marshaling research against a public-health enemy that was widely judged a failure after the hundreds of millions of dollars spent on research failed to make a dent in cancer mortality rates over the next two decades. More recently, NIH administrators have unhappy memories of a 2-year strategic planning exercise for the entire agency under former Director Bernadine Healy. The effort was roundly criticized by the community and then disavowed by the Bush Administration.

"It's going to take a while for us to reach a consensus on GPRA," says Gail Jacoby, deputy associate director of planning at the National Institute on Aging and head of a panel studying the issue. "It's a little harder for us after watching the strategic plan crash and burn."

How to judge?

The concern expressed most frequently by research managers is the difficulty of finding

nontrivial measures with which to judge research productivity. Although the law offers agencies a chance in some cases to use qualitative measures, "what OMB doesn't want are silly proxies and things that are easy to measure—number of citations, publications, and so on," says David Schindel, who directs an NSF pilot focused on its 25 science and technology centers. Instead, NSF has drawn up an array of measures for its science and technology centers that correspond to the program's three broad goals—conducting cutting-edge interdisciplinary research, training students with unique capabilities to work in all sectors, and fostering partnerships with industry.

Army Lab Uses Traditional Measures

Army Research Laboratory

Program/Indicator	FY 94	FY 95
1. Pre-eminence in science and engineering		
Number of invention disclosures	100	100
Refereed papers	275	280
% Ph.D.s on staff	23%	24%
2. Closer interaction with industry, academia, and other government labs		
New cooperative agreements (CRADAs)	26	30
Guest researchers	150	180
Postdocs	42	57

SOURCE: Excerpt from the agency's FY 95 performance plan submitted to the Office of Management and Budget

The centers program will be judged according to how many centers achieve one, two, or all three of these goals. Such an approach not only forces the centers to look beyond their scientific output toward broader measures that affect society, but it also judges them as a group. "This is a high-visibility program that is supposed to take risks," says Schindel. "That's what we're doing, and we should also be ready to accept some failures."

NSF is preparing to award a half-million-dollar contract for an outside organization to collect the data needed to measure performance; the first results will be available next fall. And the directors of the multimillion-dollar-a-year centers profess to be looking forward to the challenge. "NSF has good experience with peer review in analyzing the quality of a program, but Congress and the public want more," says Lansing Taylor, director of a biotechnology and light microscopy imaging center at Carnegie Mellon University in Pittsburgh. "The key is to combine those traditional measures with more quantifiable measures without turning it into bean counting." Adds Mike White, director of the Center for Synthesis, Growth, and Analysis of Electronic Materials at the University of Texas, Austin, "I'm comfortable being evaluated on what we are doing, but there are certain things some centers clearly can do better than other centers."

Another issue for research managers is how high to set the bar. NOAA has been grappling with that problem in drawing up goals and performance measures for a program to improve the timeliness and accuracy of short-term storm warnings, based on a combination of increased monitoring, new technology, and advances in modeling.

Kathryn Sullivan, NOAA's chief scientist, notes that the agency already collects good data on the accuracy of weather forecasts, and it is relatively easy to set goals in terms of percentage improvements. But "there's still a problem of whether you really can attain [the goals], or whether it's even feasible to target [a specific] level of improvement in accuracy in 2 years," says Sullivan. "Maybe it's very feasible, but it takes 4 years or 6 years. Or maybe it's not feasible, and you're going to have the continuing challenge of not knowing when you've done enough to convince yourself that the current technology is not capable of solving the problem." In spite of these difficulties, Sullivan says, GPRA has real value as a management tool to improve how the agency operates.

NIH officials are far more skeptical of the value of applying performance measures to biomedical research. "What if the heart institute succeeds in lowering the incidence of heart disease and, as a result of living longer, more people develop cancer?" says NCI's Cherie Nichols, a member of an interinstitute panel discussing how to implement GPRA. "What do you do when a diagnostic tool for prostate cancer results in more deaths being attributed to that disease?"

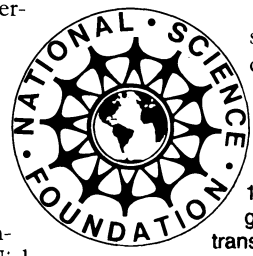
Is falling short failure?

One overriding worry for federal officials is that the goals and measures they develop may come back to haunt them. Will an agency that falls short in a particular area be penalized with budget cuts and program cancellations? "If you said you'd reach a goal in 2 years and you don't, do we increase or decrease your budget?" says NOAA's Sullivan. "Is it your fault? Is it bad management? Was it a poor experiment design, or just that much more difficult a problem than you anticipated?" NCI's Nichols concurs. "We have to be very careful about what we claim, because I think we will be held accountable. Even if we say, 'It's not a promise; it's a goal,' most people won't make that distinction."

Research managers are also concerned that the information will make it easier for OMB and Congress to compare the performances of different research agencies when budget decisions are being made. "We have to prepare against the possibility that OMB will want to use GPRA to pit agencies and programs against each other," says Gary Steinberg of the controller's office at the National Aeronautics and Space Administration. "It's a logical use of what they will be collecting."

For most agencies, the law's full impact won't be felt until the fall of 1997—one year after they have adopted performance measures—when they must incorporate those measures into a proposed budget for fiscal year 1999. Those budget submissions will certainly look far different from the traditional descriptions of programs and year-to-year budgets. "For 70 years we've had agencies preparing budgets and doing planning at the same time," says NSF's Sunley. "But the planning component was never part of the budget. Now they have to be integrated and linked to performance goals for which you can provide reasonable indicators of progress."

OMB's Walter Groszyk, who oversees some 71 pilot projects at 28 agencies being conducted under the law, acknowledges that



NSF Tries to Define Quality

National Science Foundation

1. Science and technology centers have three goals: to conduct interdisciplinary research, transfer knowledge, and graduate students with unique skills. The program will be considered "minimally effective" if 80% of the 25 centers achieve at least one goal and 50% meet two or more, and "fully successful" if 90% achieve at least one goal, 75% reach two, and 20% satisfy all three.

A. Impact of research:

Significant progress: "Most" publications appear in the field's "most respected, peer-reviewed vehicles"; Outstanding progress: Outside specialists judge several papers to be "among the most influential" in the field.

B. Collaborations:

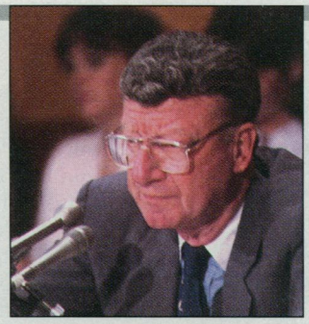
Significant progress: Such collaborations are "routine," and the center is used by "a modest number" of outside researchers; Outstanding progress: Collaboration is a "key" element, with outsiders "routinely integrated" into center-based activities.

C. Training:

Significant progress: A "majority of graduates" find productive jobs. Outstanding progress: "Nearly all graduates" become "outstanding contributors" to the work force.

SOURCE: Excerpt from the agency's FY 95 performance plan submitted to the Office of Management and Budget

research agencies probably face the most difficult challenge in complying with the new law. An analysis of the first round of performance plans submitted by the pilot projects classified about one fifth as exemplary, including NOAA's, while a similar number lacked sufficient details to be useful for measuring performance. "Put another way," Groszyk wrote his boss, OMB Director Alice Rivlin, in an August memo, "if this were 1997, little or nothing worthwhile could be salvaged from agency plans such as these." But Groszyk has not lost faith. "The pilot projects have demonstrated that measurable, quantitative goals can be set in advance," he continued. "This validates a basic underlying premise of GPRA."



RON SACHS

Performance counts. Sen. Roth (R–DE) is author of 1993 law.

That conclusion doesn't surprise Genevieve Knezo, a science policy analyst at the Congressional Research Service who has followed GPRA closely. "Is the task of complying with GPRA really as daunting as some people think? I don't think so," Knezo told participants at OSTP's November workshop. "Congress and OMB do not want lots of new data or glossy tomes full of evaluations and measures for each research project [funded]. The goal of GPRA is to ensure that agencies develop standards and processes that allow Congress to assess the performance of the nation's investment in research and to use this information in budget making. If scientists expect public support, they need to justify their claim just like any other interest group."

–Jeffrey Mervis

TAXONOMY

New Rule Could Squelch Shipments

Each year the Canadian Museum of Nature loans U.S. scientists thousands of dead specimens, from fungi to mollusks, as part of a global flow of material that helps scientists characterize new species and identify those that could pose a threat to human health or agriculture. But a proposal by the U.S. Fish and Wildlife Service (FWS) to thwart trafficking in endangered and protected species could pinch that flow and hinder the spread of knowledge of the natural world by making it impossible to ship some specimens and by raising the cost of bulk shipments. The museum has served notice, in a letter to FWS Director Mollie Beattie, that the rule would mean an "almost complete cessation of specimen loans from Canadian museums to any institution in the United States." And it's not just one country's specimens that are at stake: "The rule will stifle much of the effort in the United States to document and conserve [global] biodiversity," asserts F. Wayne King, chief herpetologist at the Florida Museum of Natural History.

The regulation is intended to strengthen enforcement of three major wildlife protection laws—the U.S. Endangered Species Act, the Migratory Bird Treaty Act, and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). To accomplish that goal, FWS would require extensive documentation of every specimen that enters or leaves the United States, dead or alive, by species name and country in which the specimen was first collected. The idea is to reduce or eliminate the multimillion-dollar illegal trade in endangered species such as parrots for sale as pets, crocodile skins for purses, and rhinoceros horns for use as aphrodisiacs.

But the proposal would also snare U.S. scientists conducting research, and more than 100 organizations and individuals wrote to FWS by last month's deadline for comments, decrying the impact of the rule on scientific activities. One problem is that many specimens are sent to U.S. institutions in order to be identified. That means prior knowledge—and documentation—of the species name is impossible. In addition, many developing countries send unidentified agricultural pests or potential disease vectors such as ticks or mosquitoes to U.S. experts to be evaluated. "This is an important and usually free service that ... would be prohibited by the regulations as proposed," says University of Kansas entomologist Charles Michener.

Scientists are also upset by FWS's intent to classify every shipment of eight or more "similar" specimens as a commercial transaction. Such a designation would subject large specimen shipments to import duties and taxes. While institutions often ship fewer than eight mammals or other large specimens at a time, King notes, "several hundred specimens might be contained in a series of minnows collected in a single seine haul or insects collected at an ultraviolet light trap." The classification of scientific transactions as commercial "simply reflects a total lack of understanding by [FWS] of the scientific enterprise," says Ross Simons, assistant provost for science at the Smithsonian Institution.

ASC

Bottled rage. Scientists say proposed rule would reduce exchange of species such as exotic snakes.

The new rules would also make U.S. scientists responsible for seeing that the material they receive in the mail contains all the requisite paperwork. "This would close down our whole operation," says Sievert Rohwer, a University of Washington ornithologist. Russian colleagues send Rohwer northern shoveler and green-winged teal ducks as part of a study on their molting and migration cycles. The birds are abundant in Russia, but happen to be listed under CITES as a protected species halfway around the world in Ghana. Shipping duck specimens requires a CITES permit, but "our collaborators look at this red tape, shrug their shoulders, and can't be bothered," says Rohwer. "Now we're going to be considered lawbreakers."

Scientists say they should be exempted from the proposed rule. Elaine Hoagland, executive director of the Association for Systematics Collections (ASC), argues that FWS could adopt a registry of research collections and scientific institutions established by the CITES governing body. By following the CITES guidelines, says Smithsonian's Simons, ASC and other organizations could accredit bona fide scientists for exemption.

But FWS officials disagree. "The government should not be in a position of determining who is a bona fide scientist," says Paul McGowan, an FWS attorney and former biologist who helped draft the rules.

Revisions to the proposed rule are expected in 6 months, and McGowan says FWS is considering "a number of solutions to facilitate the movement of scientific specimens" that stop short of a blanket exemption. In the meantime, scientists are upset at being asked to prove they are not smugglers. "There's something desperately wrong with the attitude that we're the enemy, like someone who imports massive quantities of baby green iguanas," says Harry Greene, chief herpetologist at the University of California, Berkeley's, Museum of Vertebrate Zoology. "We're honest scientists."

–Richard Stone