

December 1998

MANAGING FOR RESULTS

Measuring Program Results That Are Under Limited Federal Control



General Government Division

B-280633

December 11, 1998

The Honorable James M. Jeffords
Chairman
The Honorable Edward M. Kennedy
Ranking Minority Member
Committee on Labor and Human Resources
United States Senate

Seeking to promote improved government performance and accountability through better planning and reporting of the results of federal programs, Congress enacted the Government Performance and Results Act of 1993 (the Results Act or GPRA). This Act established a governmentwide requirement for agencies to report annually on their results in achieving their agency and program goals. Agencies can establish annual performance goals for the products and services they deliver, but they are particularly encouraged to include goals that represent outcomes, or the results of those products and services.

In the spring of 1998, agencies submitted their first annual performance plans, setting goals for fiscal year 1999. We found that many of these first performance plans faltered at the central task: developing measurable goals for the results or outcomes that their programs are intended to achieve. A common challenge faced by many federal agencies is developing goals for outcomes that are the results of phenomena outside of federal government control. Indeed, many, if not most, federal programs aim to improve some aspects of complex systems, such as the economy or the environment, or share responsibilities with other agencies for achieving their objectives, and thus face the challenge of setting goals that both are far-reaching and can be realistically affected by the programs.

To assist agencies in identifying methods for developing such goals, we conducted six case studies of how agencies were able to address the challenge of developing performance measures for outcome goals that are influenced by external factors. This report, which we prepared under our basic legislative responsibilities, discusses the strategies that these six agencies employed in setting outcome goals. Because of your interest in improving the quality of information on federal programs, we are addressing this report to you.

To find these six cases, we reviewed agency performance plans for fiscal year 1999. We selected six cases to represent a variety of programs in different agencies that addressed this challenge. We interviewed program

officials and reviewed published materials to determine answers to the following questions: (1) What strategies or techniques did they use to address this challenge? (2) What additional analytic challenges did they face, and what strategies did they use to address these challenges? (3) What special resources or circumstances, if any, were identified as important to their efforts?

Results in Brief

The six cases we studied shared the challenge of having limited control over the achievement of their intended objectives. Five of the six agencies proposed a mix of outcome goals in their annual performance plans to include far-reaching or end outcomes as well as intermediate outcomes within their more direct control. For example, one agency proposed to measure both highway fatalities (end outcome) and seat belt use (intermediate outcome). In addition, some of these agencies (1) employed a variety of analytic strategies—such as breaking out data on subgroups of clients or making statistical adjustments—to attempt to reduce the influence of external factors on their measures or (2) narrowed the scope of their measures to reflect more closely the populations served—such as employees in targeted industries.

Overall, the six agencies also employed a range of strategies to address additional challenges that arose from the particular circumstances of their programs. For example, where measures of an ultimate goal—such as prevention of a disease that takes years to develop—were unavailable, three agencies instead relied on assessing whether research-based prevention practices were in place. Three other agencies with great variability in their activities from site to site that made it difficult to set common intermediate outcomes instead relied on end outcomes as a common measure across sites. For example, while local employment assistance sites may tailor preparation activities to the needs of the clients and local labor market, these sites were all measured against clients' subsequent employment. Agencies also varied in their strategies for obtaining common data to portray their programs at the national level. Two agencies extracted common data from existing state records, such as police accident reports, while three others developed their own data collection and reporting systems, such as follow-up interviews with clients. Two agencies drew on the results of independent data sources, and one of these agencies also proposed to use national program evaluations to assess states' progress on varied intermediate outcomes.

In developing their performance goals, all of the agencies appeared to have benefited from considerable and perhaps unusual access to analytical resources and from previous experience in measuring their results. Three

programs had legislatively mandated reporting requirements; three agencies had begun strategic planning to identify their mission and long-term goals before the Results Act was enacted. In each of our cases, officials had access to research on the relationship between their programs' activities and intended results or had experience using research and evaluation in program planning. Several agency officials mentioned the importance of stakeholder involvement in the development of practical and broadly accepted performance measures. Three programs used performance information to hold local service providers accountable for results.

Background

The Results Act seeks to improve the efficiency, effectiveness, and public accountability of federal agencies as well as to improve congressional decisionmaking. It aims to do so by promoting a focus on program results and providing Congress with more objective information on the achievement of statutory objectives than was previously available. The Act outlines a series of steps whereby agencies are required to identify their goals, measure performance, and report on the degree to which those goals were met. Accordingly, executive branch agencies submitted strategic plans to the Office of Management and Budget (OMB) and Congress in September 1997 and submitted their first annual performance plans in the spring of 1998. Starting in March 2000, each agency is to submit a report comparing its performance for the previous fiscal year with the goals in its annual performance plan.

In a May 1997 report,¹ we identified the following four steps or activities in the performance measurement process to represent the analytic tasks involved in producing the documents required by the Act:

- identifying goals: specify long-term strategic goals and annual performance goals that include the outcomes of program activities,
- developing performance measures: select measures to assess programs' progress in achieving their goals or intended outcomes,
- collecting data: plan and implement the collection and validation of data on the performance measures, and
- analyzing data and reporting results: compare program performance data with the annual performance goals and report the results to agency and congressional decisionmakers.

¹ Managing for Results: Analytic Challenges in Measuring Performance (GAO/HEHS/GGD-97-138, May 30, 1997).

We reviewed agency annual performance plans submitted to Congress in the spring of 1998 on the basis of the requirements of the Act; the legislative history;² guidance contained in OMB Circular A-11, part 2; and our published guidance to evaluators and Congress on issues to consider in assessing agency performance plans.³ We issued numerous reports on the results of those individual reviews as well as a capping report summarizing the issues identified across these reviews.⁴

From our review of agencies' first performance plans, we identified a common weakness, namely, that few performance goals were outcome-oriented. In many cases, agencies faced a common challenge that we had identified in the May 1997 report: setting measurable goals for outcomes that are the result of complex systems or phenomena outside of government control.

In the previously mentioned report, we found that agencies conducting performance measurement pilot efforts often found this to be challenging because it was difficult to confidently attribute a causal connection between the program and its desired outcomes. Thus, in cases where external factors influence the program's outcomes, an examination of performance measures alone will not accurately reflect a program's performance or effectiveness. In the past, agencies have conducted systematic studies of program effectiveness, or impact evaluations, to establish the causal connection between a program's activities and its intended outcomes. To assess the net effect of a program, impact evaluations apply scientific research methods to compare program outcomes with an estimate of what would have happened in the absence of the program. Although the Results Act does not require agencies to conduct formal impact evaluations, it does require them to (1) measure progress toward their goals, (2) identify which external factors might affect such progress, and (3) explain why a goal was not met. Thus, to accurately portray program performance, it becomes important for agencies to try to control for the influence of external factors on their performance measures.

² S. Rep. No. 58, 103d Cong. 1st Sess. (1993).

³ Agencies' Annual Performance Plans Under the Results Act: An Assessment Guide to Facilitate Congressional Decisionmaking (GAO/GGD/AIMD-10.1.18, Feb. 1998) and The Results Act: An Evaluator's Guide to Assessing Agency Performance Plans (GAO/GGD-10.1.20, Apr. 1998).

⁴ Managing for Results: An Agenda to Improve the Usefulness of Agencies' Annual Performance Plans (GAO/GGD/AIMD-98-228, Sept. 8, 1998).

Scope and Methodology

To assist agencies in identifying methods or strategies for developing outcome goals in situations where goals were influenced by external factors, we conducted case studies of how six agencies were able to propose such goals in their annual performance plans. To select these cases, we examined our reviews of the departments' performance plans for examples that we noted as having set outcome goals. We then reviewed these performance plans to identify programs or agencies (below the departmental level) that had set goals for outcomes that were subject to the influence of factors outside of their programs. We next selected six cases, in concert with our teams that reviewed the agency plans, to represent a variety of strategies, program structures, and content areas. For example, while two of our selections are regulatory programs in the health and safety area, the remaining four selections represent service programs in the areas of education, job training, health and safety, and resource conservation. Also, two cases represent the direct operations of a federal agency; programs in the other cases operate through state and local agencies or the private sector. Lastly, three of the cases have their own statutory reporting requirements. Our cases consist of three individual programs and three agencies below the departmental level that proposed goals to cover more than one program. All six cases are described in the next section of this report.

To identify the analytic challenges these agencies faced and the strategies they used to address them, we analyzed the performance plans and other published materials about these programs, drawing on the analytic challenges and strategies identified in our previously mentioned report of agencies' pilot efforts. We then confirmed our understandings with federal agency officials who were involved in developing the performance plans and obtained additional information on these officials' challenges and strategies and the resources or circumstances that assisted their strategies. However, we did not independently verify the information they provided.

We conducted our work between June and September, 1998, in accordance with generally accepted government auditing standards. We requested oral comments on a draft of this report from the heads of the agencies responsible for our six cases. During October and November, 1998, we contacted officials from each agency who either said they did not have comments or responded with technical comments, which were incorporated as appropriate.

Program Descriptions

For each of the six case studies, we describe in the following paragraphs the program's or agency's mission and major activities.

Job Training Partnership Act. In the Department of Labor (DOL), the Job Training Partnership Act (JTPA) Title II programs aim to establish job training programs to assist economically disadvantaged adults and youths, and others who face significant employment barriers, to obtain (or prepare for) self-sustaining employment. DOL provides financial and technical assistance through the states to local Service Delivery Areas (SDA) to provide job training and other services designed to increase employment and earnings, develop educational and occupational skills, and decrease welfare dependency. DOL sets performance standards and measures for SDAs regarding, among other things, program clients' job retention and wage levels after leaving the program. States review and approve SDA plans for providing services, monitor program activities for compliance, and can provide (1) incentive payments to SDAs that exceed their performance standards or (2) technical assistance to SDAs that miss their targets.⁵

National Highway Traffic Safety Administration. The Department of Transportation's (DOT) National Highway Traffic Safety Administration (NHTSA) has the following as a strategic goal: reduce highway crashes, fatalities, injuries, and property losses. To achieve this, NHTSA sets safety performance standards for motor vehicle production and provides financial and technical assistance to states and local communities so that they can conduct highway safety programs that respond to local needs. In general, states set and enforce their own laws regarding highway safety. Under the State and Community Highway Safety grants, NHTSA funds projects related to driver behavior.⁶ NHTSA also conducts research and development in vehicle design and driver behavior to identify the most effective and efficient means to bring about safety improvement.

Natural Resource Conservation Service. The Department of Agriculture's (USDA) Natural Resource Conservation Service (NRCS) has the following as one of its strategic goals: a healthy and productive land that sustains food and fiber production, sustains functioning watersheds and natural systems, enhances the environment, and improves urban and rural landscapes. To achieve these outcomes, NRCS field staff, often in concert with state environmental agency staff, are to provide assistance to resource managers to help them plan, design, implement, and maintain

⁵ The recently enacted Workforce Investment Act consolidates job training programs and is to eventually replace JTPA, but it is not yet clear how the legislation will affect the performance standards systems and DOL's performance goals.

⁶ Projects related to both driver behavior and road conditions—such as speed control—could be jointly funded by NHTSA and the Federal Highway Administration.

systems to conserve, improve, and sustain natural resources and the environment. NRCS administers several conservation programs that aim to reduce soil erosion; improve air, water, and soil quality; and improve and conserve specific types of habitats, such as wetlands, croplands, and grasslands. NRCS also conducts natural resource inventories and assessments and develops conservation standards and guidelines.

Occupational Safety and Health Administration. Also in DOL, the Occupational Safety and Health Administration's (OSHA) workplace safety and health programs aim to promote safe and healthful workplaces. OSHA attempts to reduce workplace injuries, illnesses, and fatalities through developing and enforcing occupational safety and health standards, educating workers and employers about workplace hazards, and providing assistance to employers to gain compliance with those standards. OSHA directly oversees and enforces its standards and provides assistance in about one-half of the states, which cover about 60 percent of workplaces. In the other states, where OSHA has determined that their standards and enforcement capacity are at least equivalent to those of the federal program, the state agencies operate their own safety and health programs with 50-percent federal funding, and OSHA monitors their performance.

Safe Drinking Water Program. The Environmental Protection Agency's (EPA) Safe Drinking Water (SDW) Program has the following as its strategic goal: improve and maintain drinking water safety and, thereby, health protection for the 240 million Americans who get their drinking water from public water systems. EPA sets standards for drinking water filtration and disinfection processes and maximum contaminant levels and provides technical assistance and other support to the states, which have primary enforcement authority. States, in turn, are to oversee local water suppliers' implementation of federal drinking water regulations and conduct assessments of drinking water sources and potential sources of contamination. Water system operators are to routinely test their water supplies and report the results to the state agency.

Title I: Education Assistance. In the Department of Education (ED), Title I of the Elementary and Secondary Education Act aims to improve the teaching and learning of children in high-poverty schools to enable them to meet challenging academic content and performance standards. To accomplish this, ED provides technical assistance and grants to state education agencies and through them to local school districts in accordance with the number of children from low-income families. In this program, states are required to set challenging standards and student performance assessments that apply to all students and use them to assess

whether schools receiving funds are making adequate yearly progress. Schools decide how to spend their Title I resources and can combine resources from various programs to support comprehensive schoolwide reform. However, schools are encouraged to increase the amount and quality of instructional time, upgrade curriculum and instruction, provide teachers with access to professional development, and increase parental involvement.

Agencies Used a Variety of Strategies in Situations Where They Have Limited Control Over Outcomes

A common challenge faced by all six cases was having limited control over the influence that external factors have on the achievement of their strategic outcome goals. Therefore, the agencies were faced with the dilemma of whether to select (1) annual performance goals that represent the ultimate benefits of their activities to the taxpayer or (2) goals that they could reasonably expect to achieve directly and for which they could be held accountable. We identified four strategies that agencies used to address this challenge, occurring at different steps throughout the performance measurement process. Each strategy aims to reduce, if not eliminate, the influence of external factors on the agencies' outcome measures:

- selected a mix of outcome goals over which the agency has varying levels of control,
- redefined the scope of a strategic goal to focus on the more narrow range of their actual activities,
- disaggregated goals for distinct target populations for which the agency has different expectations, and
- used data on external factors to statistically adjust for their effect on the desired outcome.

A Common Strategy: Set a Mix of Goals

Five of our six cases addressed the challenge of limited control in part by selecting a mix of performance goals that included end outcomes—representing at least some of the ultimate benefits desired—and intermediate outcomes—representing conditions believed to precede or contribute to achieving the ultimate benefits. This allowed agencies to minimize the risk due to their limited control over external factors: if unexpected events prevent agencies from achieving the end outcome, they may be able to demonstrate their effectiveness through the intermediate outcome.

Table 1 gives a selection of the programs' strategic goals and outcome measures as presented in their departmental performance plans and illustrates the difference between intermediate-outcome and end-outcome goals. For example, NHTSA's mission is to reduce motor vehicle crashes

and their consequent fatalities and injuries. NHTSA believes that its performance standards for vehicles directly contribute to reducing both the incidence and severity of crashes that occur. However, occupant behavior is not as readily influenced by the federal program as are vehicle characteristics. Thus, although an increase in the use of safety belts is considered an outcome goal, it is an intermediate outcome because it is desirable not in itself but because it is believed to contribute to the ultimate goal—reducing highway-related fatalities and injuries.

Table 1: Selected Strategic Goals and Proposed Outcome Measures for the Six Federal Programs or Agencies Studied

Program or agency	Strategic goals	Intermediate outcomes	End outcomes
Job Training Partnership Act programs (JTPA)	Enhance opportunities for America’s workforce: Increase employment, earnings, and assistance. Assist youth in making the transition to work.	Percentage of youth clients employed or advancing education or job skills at program exit. Percentage of welfare clients employed at program exit.	Increase percentage of adult clients employed and average earnings level 12 weeks after program exit.
National Highway Traffic Safety Administration (NHTSA)	Promote the public health and safety by working toward the elimination of transportation-related deaths, injuries, and property damage.	Increase rate of front-seat safety belt use. Reduce number of alcohol-related fatalities.	Reduce rates of transportation-related fatalities and injuries per 100 million vehicle miles traveled.
Natural Resources Conservation Service (NRCS)	Maintain a healthy and productive land that sustains food and fiber production, sustains functioning watersheds and natural systems, enhances the environment, and improves urban and rural landscapes.	Nutrient, irrigation water, animal waste management systems, and other resource management systems applied.	Acres of cropland protected from erosion. Wetlands created or restored, and native grassland vegetation restored. Miles of conservation buffer restored.
Occupational Safety and Health Administration (OSHA)	Quality workplaces: Foster quality workplaces that are safe, healthy, and fair. Reduce workplace injuries, illnesses, and fatalities.	Reduce levels of lead and silica exposure.	Reduce three of the most prevalent workplace injuries/illnesses. Reduce injuries/illnesses in five high-hazard industries. Reduce injuries/illnesses in workplaces where OSHA intervened. Reduce construction fatalities.
Safe Drinking Water (SDW) Program	Improve and maintain drinking water safety, and thereby provide health protection for the 240 million Americans who get their drinking water from public water systems.	Increase percentage of population having access to water meeting health-based standards.	(None reported.)
Title 1: Education Assistance program (Title I)	At-risk students improve their achievement to meet challenging academic content and performance standards.	States adopt challenging performance standards. Schools improve teacher training and curriculum and instruction, and extend learning time.	Increase mathematics and reading test scores among children in high-poverty schools.

Note: This table may not include all proposed performance measures but ones that reflect the mix of intermediate outcomes or end outcomes.

Source: GAO analysis of agency performance plans.

In other cases, because these agencies often had multiple goals, the intermediate outcomes did not necessarily lead to a proposed end-outcome goal, but rather to an unmeasured end outcome. For example, the JTPA program sets its goal for youth clients as employment or enrollment in advanced training at the time of program departure. Since obtaining these placements is part of the program’s responsibilities, we considered this to be an intermediate outcome goal. However, the JTPA program did not propose a parallel measure for the potential end outcome of the youths’ obtaining self-sustaining employment.

Narrow the Goal’s Scope to the Program’s Span of Influence

Table 2 displays the strategies used by the six agencies for both the common challenge involving external factors and additional challenges categorized by the step or activity of the performance measurement process in which the strategy was employed.

Table 2: Strategies Used by the Six Agencies Reviewed to Address Challenges Throughout the Performance Measurement Process

Challenge	Performance measurement activity			
	Identify goals	Develop measures	Collect data	Analyze results
Limited control over intended outcomes	Narrow scope to span of influence (JTPA, NRCS, SDW, OSHA).	Set mixture of intermediate-outcome and end-outcome goals (Title I, NRCS, JTPA, NHTSA, OSHA).		Disaggregate data for groups with different performance expectations (JTPA, NRCS, OSHA). Statistically adjust to control for external factors (JTPA, NHTSA).
Multiple goals	Set separate goals (Title I, NRCS, NHTSA, JTPA, SDW).	Combine goals into a single measure (JTPA).		
End outcomes take years to develop	Focus on intermediate goals that research demonstrates are linked to ultimate goal (NRCS, OSHA, SDW).			
Variability in local program activities	Rely on measures of common end outcomes (JTPA, NHTSA, OSHA).			
Variability and incompatibility of data		Adopt results of independent data sources (OSHA, Title I).	Extract state data (NHTSA, SDW). Institute own data-collection requirements (JTPA, OSHA, NRCS).	Summarize evaluation results (Title I).
Potential data collection burden		Replace detailed reports of project activities with reports of projects completed (NRCS).	Institute sampling (JTPA, Title I). Target a few grade levels (Title I).	

Note: Absence of table entry does not imply that a strategy does not exist; only that one was not identified in our review.

Source: GAO analysis of documents and interviews with federal officials.

Some programs have strategic goals that imply they have more extensive or a broader range of activities than they in fact do. In such cases, agencies identifying goals may choose to define a narrower scope for the performance goal, which, from the agency's perspective, is a more realistic target. OSHA's workplace safety and health programs provide several examples of how this can be done.

Due to the large number of workplaces in the country, it is impractical for OSHA to routinely perform health and safety inspections in all workplaces. Instead, program officials indicated that they target their activities to where they see the greatest problems--those industries and occupations with the highest rates of fatalities, injuries, or illnesses. In addition, the program only conducts major interventions involving compliance assistance in a limited number of workplaces each year. Thus, OSHA does not realistically expect to be able to dramatically reduce the number of all workplace injuries and illnesses each year. Instead, the agency sets a series of performance goals that reflect these different levels of influence. For fiscal year 1999, the department plan proposes a 3-percent reduction in three of the most prevalent injuries and illnesses and a 3-percent reduction in injuries and illness in five "high-hazard" industries. Moreover, where OSHA can be more confident of having an effect, such as where they have launched an intervention, agency officials propose a 20-percent reduction in subsequent injuries and illnesses in those worksites.

Many federal programs operate through the actions of state agencies, raising questions about whether and how federal programs should be held accountable for the actions of others. In this case, OSHA chose to set goals only for workplaces in states in which the federal program has primary enforcement authority. Agency officials explained that they were not comfortable with being held responsible for the actions of others. However, they noted that states are required to prepare and submit strategic and annual performance plans consistent with OSHA's strategic plan, and that states' results will be included in OSHA's performance reports. Four of the other cases we reviewed also operate through state or local agencies, but their programs do not have a split like OSHA's where almost as many states have primary authority as do not. Each of these four cases viewed its program as the sum of the activities of each responsible party and set performance goals to represent its outcomes as a whole.

Set Separate Goals for Populations With Different Performance Expectations

Some programs serve distinct groups for which performance expectations differ, but these programs have little control over changes in the relative size of those groups. Yet, with no actual change in program effectiveness, an increase in the size of the poorest performing group could drag down an

indicator of overall performance and make it look as if the program had become less effective. To avoid this problem, programs can track separate disaggregated performance goals for these groups. The previously discussed OSHA strategy of calculating injury and illness rates for specific high-hazard industries is an example of this. Calculating separate rates by industry helps control for the possibility that an increase in an overall injury rate over time might simply reflect an increase in employment in hazardous industries. It also allows OSHA to track the impact of its programs on targeted industries.

In another example, the JTPA program has less experience with long-term welfare recipients than with its other adult clients on whom it has been collecting post-program job retention data for some time. Therefore, JTPA's performance plan noted that its performance target for its welfare clients was provisional and subject to change. Additionally, because youth participants (younger than age 22) are served by a separately authorized program whose goal is not employment, per se, but transition to employment, the agency set a separate performance goal of these youths either being employed or obtaining advanced education or job skills at program completion.

Statistically Adjust for the Effects of External Factors

Carrying the previous strategies a step further, if the role of external factors is reasonably well understood and data are available, explicit statistical adjustments can be made for their effects. For example, in analyzing results, NHTSA uses the ratio of fatalities per vehicle mile driven to control for the simple fact that if more miles are driven, then more crashes are likely to result. Ratios and rates can be very useful for measuring outcomes that are related to population size. Since the national population grows every year, any phenomena related to it (e.g., the number of cars on the road and the number of crashes) will generally grow along with it, even though individuals may not be more likely to drive or to have an accident.

As part of its performance standards system, the JTPA program developed statistical regression models predicting client job retention and wage levels that are based on socioeconomic factors, such as client characteristics and economic conditions. These models are used in generating standards for local SDAs by applying figures from their local labor markets and caseloads to the appropriate model. The goal of these adjustment models is to set realistic goals for each SDA and not penalize them for differences among them in local conditions or the type of participants they serve. However, it should be noted that in other programs, adjusting expectations on the basis of clients' socioeconomic

status has been criticized as perpetuating lowered expectations for the economically disadvantaged.

Programs Addressed Additional Challenges Reflecting Their Specific Circumstances

The six agencies we studied also addressed additional challenges in developing outcome-oriented performance measures that reflected the particular circumstances of each program. We described the strategies that agencies used to overcome these various situations in table 2. The additional challenges were as follows:

- A program may have to balance multiple goals or key dimensions of a goal.
- The effect of program activities may take many years to observe.
- Approaches and activities may vary extensively around the country.
- Data collected by others may vary so extensively that they cannot be readily aggregated to provide a national picture.
- Obtaining comprehensive data on program performance may not be practical on an annual basis.

Multiple Goals Can be Measured Separately or Combined

In operationalizing a broad strategic goal into a measurable performance goal, an agency may find that it needs separate goals to capture the key dimensions of intended performance. One strategy for this challenge was to set separate outcome goals that may more completely reflect the intended results of the program. For example, the Title I program translated “academic achievement” into separate outcome goals for mathematics and reading achievement, as well as setting separate goals for particular school reforms, such as increased professional development for teachers. Similarly, NRCS created separate goals for restoration of each major type of habitat of concern (i.e., croplands, watersheds, wetlands, and grazing lands). In other cases, agencies set separate goals for areas of special interest, such as NHTSA’s separate goal for reducing alcohol-related highway fatalities.

Other agencies found that they could combine outcomes into a single measure. In the JTPA youth program, the goal is preparation for transition to work, which can be achieved through obtaining either employment or advanced education. Thus, its primary measure for the youth program is whether youths are either employed or enrolled in some post-secondary education on program completion. In the adult program, however, the standard is employment and earnings at or above a specific level to reflect the strategic goal of “self-sustaining employment” and to discourage rapid placement of clients in “dead-end” jobs. The SDW program rolls up many different standards into a single measure, the population served by systems meeting health-based standards.

Research-Based Prevention Can Provide Proxies for End Goals

Since some outcomes may take years to develop, a long lag-time may be required to see the end outcome, making it impossible to attribute the results observed to the previous 1 year's or even 2 years' activity and permitting other factors to intervene between activities and results. To deal with this analytic problem, a few agencies relied on previous research to establish links between their activities and intermediate outcomes and from those to the desired end outcome.

For the SDW program, the end goal is to protect the population from illnesses caused by water-borne contaminants, but some illnesses (such as cancer) may take many years to develop after exposure to the contaminant. Therefore, safe water is defined as water having met all standards for maximum levels of specific contaminants or required treatment techniques, because the standards were based on research establishing the health risks associated with specific levels of exposure.

Similarly, because it takes a long time to improve the quality of soil and watersheds, NRCS relies on its knowledge of effective conservation practices developed through extensive research on soil and water management, conducted by NRCS and USDA's Agricultural Research Service. These practices are spelled out in technical guides, along with criteria for whether the practices have been fully implemented. The wide acceptance of the effectiveness of these practices permits NRCS to measure progress toward the end outcome of land improvement with an intermediate outcome of the number of acres where these practices were applied.

Focus on End Outcomes Can Provide Commonality Across Diverse Program Approaches

Variability in local activities in some programs can be so extensive that there is little commonality in intermediate outcomes and, thus, little in common to measure across providers or sites to portray the "national" program. In the cases we reviewed, agencies in this situation found that the measures most common to all participants or activities were end outcomes.

For example, NHTSA funds state traffic safety programs that can target, among other things, speeding or motorcycle safety and can be implemented differently in each state. So, instead of trying to measure a reduction in speeding in one state and an increase in motorcycle safety in another, NHTSA chose the end outcomes of the rates of crashes, fatalities, and injuries.

Similarly, in the JTPA program, different sites may serve clients with different levels of job experience and may have different types of local

jobs and industries for which to prepare clients; therefore, service provision as well as program approach varies across sites. Nevertheless, all sites are to be measured against their end outcome, which is subsequent employment and wage levels for adult clients. OSHA also is concerned with different health and safety hazards in different industries, leaving little to combine except the “bottom-line” numbers of injuries, illnesses, and fatalities.

Programs Can Standardize Their Own Data Collection or Draw on Independent Sources

Programs in two of our six cases (the JTPA and SDW programs) had statutory reporting requirements and could provide common performance data from sites across the country. To fill information gaps or to obtain data that were not readily available to them, other agencies developed or adopted a standardized data collection system. For example, to obtain consistent and comparable nationwide crash information, NHTSA arranged for standardized crash data to be collected from state police and other records. NRCS collects selected data from its own site technical assistance records.

Some agencies instituted their own data collection to fill information gaps. Although the Bureau of Labor Statistics (BLS) provides OSHA with aggregate survey data on injuries and illnesses by selected industries, for privacy reasons, BLS would not disaggregate the data to disclose the identity of individual employers. Therefore, OSHA conducts its own follow-up data collection at its intervention worksites to obtain more tailored information with which to assess the effectiveness of its intervention activities.

To obtain data that were not readily available to them from routine program operations, two agencies adopted an existing standardized data collection system created outside of the program. As previously mentioned, OSHA used the results of the BLS survey of workplace injuries and illnesses to help establish its baselines. In addition, the Title I program is specifically precluded from requiring states to use any specific performance standard or test but does require states to set their own academic standards and select their own assessment instruments, so that standards can be integrated with their curricula. However, since different states may set “proficiency” standards at very different levels of achievement and use very different testing instruments, it would be extremely difficult to combine their results into a national figure. Therefore, ED adopted the results of an independent testing program, the National Assessment of Educational Progress (NAEP), which is a set of nationally standardized tests given to a representative sample of students across the nation every 2 years.

Another strategy was to conduct national evaluations to combine or synthesize program results. Because the Title I program is intended to be integrated into each state's individual plan for school reform, ED has developed a complex set of studies to constitute its legislatively mandated "National Assessment of Title I." In the National Assessment, ED proposes to deploy special evaluation studies to synthesize information from national surveys with information from the states on the progress of their own school reform initiatives and their assessments of school and student progress. ED proposes to draw on these studies to provide measures of state progress on intermediate school improvement outcomes.

Statistical Sampling and Summaries Reduce Data Collection Burden

Because the time and cost of collecting comprehensive data can be burdensome, agencies appear to be selective in what data are collected and how they are collected.

One approach is to institute sampling. The JTPA program requires follow-up interviews with a sample of program clients 12 weeks after they leave the program; however, in areas with large caseloads, not all clients need to be interviewed to obtain statistically accurate estimates. Regarding the Title I program, NAEP uses sampling at three levels. First, only children at three grade levels, 4th, 8th, and 12th grades, take the tests each time they are given. Second, only a random sample large enough to provide reliable estimates at the state level, not all children in those grades, is tested. Third, blocks of test items are systematically varied among testing booklets to obtain results on a large number of items without each student's having to answer all of the items.

A second strategy is to abstract from agency field office records only the data that agency headquarters needs. An NRCS official described this strategy as a continuing effort; traditionally, local program staff completed detailed diaries of the technical assistance they provided (such as number of consultations or number of conservation plans reviewed or revised) and reported activities in detail to headquarters. The official indicated that over time, the agency had reduced its reporting requirements and now focuses on higher level measures that represent "change on the land," such as the number of soil management systems completed.

Previous Efforts Shaped Programs' Strategies for Measuring Results

Considerable—and perhaps unusual—analytic resources and experience in using performance information and other favorable circumstances (congressional interest and other stakeholder involvement) appeared to facilitate these agencies' recent efforts to develop measures of their programs' results for fiscal year 1999.

Agencies Had Previous Experience Collecting Data on Program Results

Most of the agencies studied already had research and evaluation studies linking their activities to intended program outcomes when they began to develop their performance plans under GPRA. Although our recent survey identified limited federal resources for formal studies of program results, that is, program evaluations, many of the programs we studied had access to analytic resources and a few worked with well-established, active program evaluation units.⁷

In some instances, Congress mandated results-oriented program assessment many years before the enactment of GPRA. As part of the regulatory process, local water suppliers in the SDW program are required to report their water test results to their state agencies to demonstrate compliance. Performance measures or standards were mandated in the JTPA program since its enactment in 1982 to create accountability for local SDAs. In addition, the legislation required a national evaluation of the JTPA program's effectiveness. Since the 1970s, the Title I (and the prior Chapter 1) program has required tests of student achievement to measure the program's progress in assisting disadvantaged students. In addition, Congress mandated comprehensive national assessments of both the Title I and Chapter 1 programs. For the current National Assessment of Title I, ED's Planning and Evaluation Service plans to coordinate a large number of studies to answer diverse questions about the progress of the 1994 reforms.

Other agencies, such as NHTSA, NRCS, and OSHA, have had access to or have been collecting at least some basic data regarding program results over many years. Data concerning the circumstances surrounding fatal vehicle crashes were available from the mid-1970s, and, over the years, NHTSA developed a way to extract consistent data from states' police records. NHTSA also funds evaluations of state and local traffic safety programs. Both the Soil Conservation Service (SCS), NRCS' predecessor, and USDA's Agricultural Research Service have conducted extensive research on soil and water management, which an NRCS official indicated provides the basis for the guidance on conservation practices in their technical guides.⁸ In addition, NRCS conducted Natural Resource Inventories in 1992 and 1997 that established baselines for some natural resource and habitat conditions and to target program efforts. OSHA

⁷ Program Evaluation: Agencies Challenged by New Demand for Information on Program Results (GAO/GGD-98-53, Apr. 24, 1998).

⁸ NRCS was created in 1994 USDA-reorganization legislation by merging the SCS and several of the conservation cost-sharing programs of the Agricultural Stabilization and Conservation Service.

officials reported that BLS has been conducting employer surveys and sharing selected data with OSHA for many years.

Stakeholders' Involvement Helped in Selecting Measures of Program Outcomes

Agency officials noted that, where there were numerous parties with potentially different interests—such as program managers, third-party service providers, and customers—the agency found that involvement of these interested outside parties contributed to defining and measuring program outcomes. The advisory board for the National Assessment of Chapter 1—consisting of congressional and academic stakeholders as well as state and local education practitioners—was considered helpful by an agency official to ensuring both the credibility of the agencies' evaluation results and acceptance of their recommendations. In addition, it was the independent Advisory Committee on Testing (part of the National Assessment of Chapter 1) that recommended what became a major shift in strategy for measuring the program's results: using NAEP to provide information on national progress and state-adopted tests to assess school progress. JTPA program officials noted that their performance measurement system has been fine-tuned over the years through the cooperation of local program managers and other stakeholders. Moreover, the JTPA program has a regular process of negotiating SDA performance standards with the states every 2 years. OSHA officials also reported stakeholder consultation in the selection of performance measures.

In addition, some agencies began strategic planning under government reform initiatives before the enactment of GPRA. SCS had developed strategic plans that included agency goals and objectives since the 1970s. In 1992, EPA started a long-range, goal-setting initiative to clarify its mission and responsibilities, called the National Environmental Goals Project. EPA described having extensive stakeholder participation, including federal and state agency review and public hearings on a draft report to target environmental problems of greatest concern to citizens or that posed the greatest risk. EPA officials also noted the helpfulness of their recent participation in an intergovernmental task force on water quality with public agencies and external environmental groups to develop measures of water quality. NHTSA reported that it also began strategic planning in 1992 as part of a DOT-wide effort leading to a DOT Strategic Plan in 1994. NHTSA worked with a broad range of organizations representing the interests of the traveling public, as well as, among others, motor vehicle manufacturers, the insurance industry, state highway safety offices, and the business community, to develop agency goals.

Some Agencies Used Program Performance Information in Managing Their Programs

Some agencies used information on program results to assess past and future federal policies, while other agencies used such information to hold local service providers accountable for results. NHTSA uses the Fatality Analysis Reporting System and other data to conduct research to identify the key factors that affect crash incidence and the seriousness of consequences. NHTSA uses this research to plan and evaluate traffic safety programs and to investigate new areas for regulation. Concern about contamination of drinking water supplies coming from their source waters led to the recent SDW requirement to assess source waters for vulnerability to contaminants. EPA officials noted that measuring the size of the population served by a water supplier helped them focus on where compliance problems would have the largest potential population effects. The results of the National Assessment of Title I's predecessor program, Chapter 1, led to a major restructuring of the program and its assessment approach. Specifically, the National Assessment final report recommended encouraging performance standards for schools be tied to their curricula; focusing program efforts on improving schools; giving states flexibility in return for accountability for student performance; and aligning assessment of the students served with standards that apply to all students in the state.

In the SDW program, water testing is designed to identify the need for corrective action at the local level, such as disinfectant treatments, to bring the water quality up to standard. EPA officials noted that public disclosure of source water quality results might also mobilize communities to improve protection efforts. Both the JTPA and Title I programs designed their performance measures to hold local providers accountable for results. JTPA gives states the ability to give incentive payments to local SDAs that exceed their performance standards and technical assistance to those who fail to meet the standards. Under the revised Title I program, one purpose for measuring poverty schools' progress toward achieving statewide performance standards is to hold them accountable to parents and the community as well as to ensure high standards and expectations of the students served. Perhaps more important, some school districts around the country have used such performance data to diagnose low performance and attack specific problems with concrete solutions, such as changes in instructional practice.⁹

We are sending copies of this report to the Chairmen and Ranking Minority Members of the Senate Committee on Governmental Affairs and the House Committee on Government Reform and Oversight, the Director of OMB,

⁹ Turning Around Low-Performing Schools: A Guide for State and Local Leaders. U.S. Department of Education, Washington, D.C., May 1998.

and other interested parties. We will also make copies available to others on request.

Please address any questions to me or Stephanie Shipman, Assistant Director, at (202) 512-7997. Another major contributor to this report was Elaine Vaurio, Project Manager.



Susan S. Westin
Associate Director, Advanced Studies
and Evaluation Methodology

Bibliography

Barnow, Burt. "The Effects of Performance Standards on State and Local Programs," Evaluating Welfare and Training Programs, Charles Manski and Irwin Garfinkel (eds.). Cambridge and London: Harvard University Press, 1992.

National Highway Traffic Safety Administration. The National Highway Traffic Safety Administration FY 1999 Performance Plan. Washington, D.C.: Jan. 1998.

U.S. Department of Agriculture. FY 1999 Annual Performance Plan. Washington, D.C.: 1998.

———. USDA Strategic Plan: 1997-2002. Washington, D.C.: Sept. 1997.

U.S. Department of Education. U.S. Department of Education Annual Performance Plan: FY 1999, Vols. 1 and 2. Washington, D.C.: Feb. 27 and Feb. 25, 1998.

———. Reinventing Chapter 1: The Current Chapter 1 Program and New Directions. Washington, D.C.: Feb. 1993.

———. Reinforcing the Promise, Reforming the Paradigm. Washington, D.C.: May 1993.

———. Turning Around Low-Performing Schools: A Guide for State and Local Leaders. Washington, D.C.: May 1998.

U.S. Department of Labor. FY 1999 Performance Plans for Committees on Appropriations. Washington, D.C.: Feb. 1998.

———. "Strategic Plan: Occupational Safety and Health Administration, United States Department of Labor: FY 1997 - FY 2002." Washington, D.C.: Sept. 30, 1997.

U.S. Department of Transportation. "DOT Performance Plan FY 1999." Washington, D.C.: Feb. 1998.

———. National Highway Traffic Safety Administration. "Strategic Plan," DOT HS 808 181. Washington, D.C.: Nov. 1994.

U.S. Environmental Protection Agency. Fiscal Year 1999 Justification of Appropriation Estimates for the Committees on Appropriations (EPA-205-R-98). Washington, D.C.: Feb. 1998.

Bibliography

———, Office of Water. Drinking Water Standards for Regulated Contaminants. Internet address: <http://www.epa.gov/OGWDW/source/therule.html>, revised June 15, 1998.

Related GAO Products

Managing for Results: An Agenda to Improve the Usefulness of Agencies' Annual Performance Plans (GAO/GGD/AIMD-98-228, Sept. 8, 1998).

The Results Act: Assessment of the Governmentwide Performance Plan for Fiscal Year 1999 (GAO/AIMD/GGD-98-159, Sept. 8, 1998).

Grant Programs: Design Features Shape Flexibility, Accountability, and Performance Information (GAO/GGD-98-137, June 22, 1998).

Results Act: Observations on the U.S. Department of Agriculture's Annual Performance Plan for Fiscal Year 1999 (GAO/RCED-98-212R, June 11, 1998).

Results Act: Observations on the Department of Education's Fiscal Year 1999 Annual Performance (GAO/HEHS-98-172R, June 8, 1998).

Results Act: Observations on Labor's Fiscal Year 1999 Performance Plan (GAO/HEHS-98-175R, June 4, 1998).

Results Act: Observations on the Department of Transportation's Annual Performance Plan for Fiscal Year 1999 (GAO/RCED-98-180R, May 12, 1998).

Results Act: EPA's Annual Performance Plan for Fiscal Year 1999 (GAO/RCED-98-166R, Apr. 28, 1998).

Program Evaluation: Agencies Challenged by New Demand for Information on Program Results (GAO/GGD-98-53, Apr. 24, 1998).

The Results Act: An Evaluator's Guide to Assessing Agency Performance Plans (GAO/GGD-10.1.20, Apr. 1998).

Agencies' Annual Performance Plans Under the Results Act: An Assessment Guide to Facilitate Congressional Decisionmaking (GAO/GGD/AIMD-10.1.18, Feb. 1998).

Results Act: Information on Mission, Goals, and Measures Developed by FHWA, FTA, and NHTSA (GAO/RCED-98-34R, Nov. 14, 1997).

Managing for Results: Regulatory Agencies Identified Significant Barriers to Focusing on Results (GAO/GGD-97-83, June 24, 1997).

Related GAO Products

Managing for Results: Analytic Challenges in Measuring Performance
(GAO/HHS/GGD-97-138, May 30, 1997).

Program Evaluation: Improving the Flow of Information to the Congress
(GAO/PEMD-95-1, Jan. 30, 1997).

Executive Guide: Effectively Implementing the Government Performance
and Results Act (GAO/GGD-96-118, June 1996).

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are \$2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. VISA and MasterCard credit cards are accepted, also. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Order by mail:

**U.S. General Accounting Office
P.O. Box 37050
Washington, DC 20013**

or visit:

**Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC**

Orders may also be placed by calling (202) 512-6000 or by using fax number (301) 258-4066, or TDD (301) 413-0006.

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (202) 512-6000 using a touch-tone phone. A recorded menu will provide information on how to obtain these lists.

For information on how to access GAO reports on the INTERNET, send e-mail message with "info" in the body to:

info@www.gao.gov

or visit GAO's World Wide Web Home Page at:

<http://www.gao.gov>

**United States
General Accounting Office
Washington, D.C. 20548-0001**

**Bulk Rate
Postage & Fees Paid
GAO
Permit No. G100**

**Official Business
Penalty for Private Use \$300**

Address Correction Requested

