DECEMBER 2003

Strategies for Using State Information: Measuring and Improving Program Performance

Shelley H. Metzenbaum
Visiting Professor
Maryland School of Public Affairs
University of Maryland

IBM Center for
The Business of Government

Managing for Performance and Results Series

MANAGING FOR PERFORMANCE AND RESULTS SERIES

Strategies for Using State Information: Measuring and Improving Program Performance

Shelley H. Metzenbaum
Visiting Professor
Maryland School of Public Affairs
University of Maryland

December 2003

TABLE OF CONTENTS

FOREWORD

December 2003

On behalf of the IBM Center for The Business of Government, we are pleased to present this report, "Strategies for Using State Information: Measuring and Improving Program Performance," by Shelley Metzenbaum.

The Government Performance and Results Act celebrated its 10th anniversary this year. A major issue debated during the drafting of this law was how to treat federal agencies that relied heavily on state and local partners in implementing grant programs. How could federal agencies set goals, measures, and performance targets—and be held accountable by Congress for meeting them—in policy areas where the federal government is only one of a range of players, and sometimes not even the dominant player?

In fact, in some cases federal laws actually prohibit agencies from setting or collecting performance measures (such as in many of the block grants) and in some cases federal agencies are constrained by law from taking action to improve state performance (such as mandating state motorcycle helmet laws).

So what have federal agencies done in response to what seems an insoluble dilemma? This report examines the approach and strategies taken by several federal agencies in three classic intergovernmental arenas—environment, transportation, and education. And it offers a series of recommendations to various actors as to how they can improve performance in the continuing evolution of how to manage for results in a multi-player environment.

Paul Lawrence
Partner-in-Charge
IBM Center for The Business of Government
paul.lawrence@us.ibm.com

John M. Kamensky
Senior Fellow
IBM Center for The Business of Government
john.kamensky@us.ibm.com

EXECUTIVE SUMMARY

From the beginnings of U.S. history, debate has raged about the appropriate roles of federal versus state and local governments. The debate never stops. By constitutional intent, multiple levels of government with vaguely defined boundaries of responsibility check the powers of each other, protecting citizens.

This study is intended to help federal managers understand how, given the inevitable tensions of intergovernmental relations, they can best use performance goals and measures in working with state and local governments to deliver improved results to the public. The study also seeks to identify constructive roles that states individually and through their national organizations, Congress, and non-governmental organizations (NGOs) can play to advance the use of state performance information to enhance social outcomes.

Most federal agencies have long used goals and measures in their work with states and localities. Developments in information technology, which have dramatically reduced the costs of gathering, organizing, analyzing, and disseminating information, suggest that it is timely to assess how past organizational practices should be changed to take advantage of these developments. Further, the adoption of the Government Performance and Results Act of 1993 (GPRA), which requires every federal agency to set goals and report performance toward them annually, places the issue squarely on the agenda of agencies that depend heavily on state and local governments to accomplish their objectives.

This study does not attempt to be conclusive regarding the methods agencies should adopt. Instead, its intent is to begin a series of important conversations:

- Federal agencies need to ask and answer questions about how to use performance goals and measures constructively with the states in the context of their own history, organizational culture, opportunities, and technology. Addressing these questions directly holds great promise for program improvements.
- The U.S. Congress, federal agencies, state governments, localities, state government associations, and non-governmental organizations need to think more explicitly about the best ways federal agencies can use performance goals and performance information to work with states and local governments to deliver improved results to the public.
- The Office of Management and Budget (OMB) and Congress need to take a more aggressive role supporting cross-agency learning on this subject.
- The work of the General Accounting Office (GAO) and the National Academy of Public Administration to facilitate interagency exchange on performance management needs to continue and be expanded in order to direct more attention to the critical challenge of integrating state and local performance information into federal performance management efforts.

Case Studies

This report examines how four federal agencies in three departments have grappled with the challenge of integrating state and local performance with federal performance-management efforts. Based on these agencies' experiences, it concludes with findings and recommendations for Congress, OMB, federal agencies, states, and nonprofits to better balance the effective use of three strategies—measurement, mandate, and money—to achieve performance outcomes.

Environmental Protection Agency: Shifting from Activities to Outcomes

The report begins with a look at an agreement crafted between the U.S. Environmental Protection Agency (EPA) and the states, called the National Environmental Performance Partnership System (NEPPS). NEPPS was adopted as an overlay to over a dozen separate federal environmental laws, many of which establish specific environmental goals as well as measurement requirements for states. Over time, EPA's management of these laws, with the exception of the mandate that all states attain national ambient air quality standards, had evolved in a manner that emphasized processes over environmental performance. NEPPS was embraced by EPA and state environmental agency leaders as a way to make clear that, instead of EPA's historical emphasis on assuring state completion of a negotiated number of explicitly specified activities, the federal agency could use environmental progress and compliance outcomes as the dominant criteria for program accountability. Further, state leaders were encouraged to do the same in their own program management.

Since the adoption of the NEPPS agreement, the states and EPA have made significant progress. They jointly adopted a policy encouraging interested states to set their own performance goals, agreed on measures for assessing state performance, secured a legislative change allowing states to combine federal funds to meet state problems, revised grant regulations in support of NEPPS, and greatly increased public access to information about state environmental performance and their plans to improve it. In addition, several states and a few EPA regions have begun to make significant changes in the way their programs operate to heighten attention to priority environmental problems and the effect of agency actions on environmental outcomes. At the same time, many challenges remain. EPA and the states need to sort out how to organize and analyze state performance information to make it more useful to EPA, the states, and the public. Further, more state leaders need to step forward to take advantage of the flexibilities the NEPPS framework affords to realign their agency activities with environmental priorities in the states.

Department of Transportation: Sharing Information and Balancing Mandates

The report next turns its attention to two agencies of the U.S. Department of Transportation (DOT), the Federal Highway Administration (FHWA) and the National Highway Traffic Safety Administration (NHTSA). Both have established themselves as expert resources for state and local governments. They collect and disseminate written materials on state practices and progress. In addition, they identify more effective practices worthy of replication. Compilation of state information in an easy-to-find and easy-to-use format; analysis tailored to meet the needs of specific audiences—especially the states and others whose actions directly affect the rate of progress; problem and success identification; aggressive packaging and dissemination of the raw information, analysis, and materials supporting programs demonstrated to be effective; ongoing evaluation at multiple levels; and co-ownership of decision-making responsibilities characterize the successful work of FHWA and NHTSA with the states.

Despite a long, strong, and successful history of using performance information for mutual benefit, intergovernmental tensions still test the relationship between these two DOT agencies and the states. Mandates in 1991 for state adoption of management systems using state performance measures and DOT's initial efforts to implement the Government Performance and Results Act of 1993 ignited state concerns about federal encroachment on state independence. With regard to state adoption of management systems, although some of the mandates for state adoption were subsequently withdrawn, they nonetheless changed state behavior. With regard to GPRA, states were concerned when DOT talked about possibly linking federal

funding to state performance relative to outcome goals selected by DOT. The states resisted. They resented DOT's selection of goals without state consultation, and were wary of being held accountable for outcomes that states could not completely control. Indeed, the states questioned whether it was appropriate for the federal government to have an outcome-focused goal such as safety, because states and localities, not DOT, managed the bulk of the field activities that directly affected safety levels. The states preferred to have DOT measure its own performance relative to what federal agencies directly controlled, such as speed in getting federal funds to the states or reviewing state projects. Ultimately, DOT stuck with outcome goals, such as safety and mobility, to guide its work with the states, but opted not to link funding to performance levels. At the same time, the states' association of highway and transit officials adopted DOT's priority goal, safety, as its own.

Department of Education: Federal Report Cards Motivate and Analysis Illuminates

In the U.S. Department of Education (ED), the department's release of the "Wall Chart" in 1984 and, more recently, reports released by a nonprofit association using data assembled with ED funding illustrate the power of comparative measurement across jurisdictions whether within a state, among states, or internationally. The first example demonstrates how comparative measurement can motivate improved performance. It also underscores the power and possibility of state political leaders collectively endorsing comparative measurement after years of opposition. The second underscores the value of the federal government gathering and organizing state performance data. It shows how, just by gathering and organizing information and making it easy to use, the federal government can stimulate external analyses to find effective government programs worthy of replication.

Findings

Finding 1: Analyzing Measures Adds Value

By their actions and the way they use performance goals and measures, federal agencies can create an environment that continually harvests and re-sows the lessons of state experience. Simply gathering data from states about outcomes, activities, and inputs, and then organizing and making them readily available for everyone to study adds value:

- It reduces search costs for states regarding what other states are doing, for public interest groups allowing them to direct their efforts more precisely to areas needing attention, and for consultants developing information-based products for the states.
- It enables those served or regulated by government programs to assess program fairness and select among programs when options are available. Federal failure to carry out these tasks lessens the potential benefit, both motivational and informational, of state performance data.

Analyzing state performance information, after it is gathered, adds further value because it identifies successes and problems, and then triggers targeted follow-up inquiries to understand causes of problems, contributors to successes, and unsuccessful interventions not worth doing. Useful analysis can take many forms, simple and complex. Simpler forms of analysis, such as the identification of strong performers, identification of weak performers, and organization of performance data by category, trigger focused follow-up questions that lead to valuable insights for improving performance. More sophisticated analyses, including diagnostic analysis to find key contributors to problems and controlled studies, can isolate the specific variables that influence performance levels. Audience-tailored analysis of performance measurement increases the prospect that it will be used and supported.

Finding 2: Federally Mandated Goals Work, but Can Be Problematic

While often politically controversial at the state level, federally mandated goals for states can be powerful motivators when linked to the promise of significant rewards or the threat of significant penalties. The promise of a reward or threat of serious penalties linked to goal attainment can add to the motivational value of measures as long as those being measured do not feel so strongly threatened that they try to have the goals repealed or destroy the measurement system, either by dismantling it or by undermining it with inaccurate and untimely measurement submission.

Recommendations

Recommendation 1: Collect, Organize, and Make Information Readily Available

To encourage use and analysis of the information, federal agencies that depend on state and local governments for the accomplishment of their goals should annually compile state information into a single compendium that is easy to find and interpret. This information should be made available in print form and accessible online through a single portal, produced on a regular schedule, and broadly disseminated. Federal agencies could use their annual GPRA performance reports for this purpose. At a minimum, they should cross-reference where relevant state data can be found in the annual GPRA performance reports.

Recommendation 2: Create Robust Measurement Systems

Goals and measures together have the greatest motivational and informational potential. While measures can effectively drive performance without goals (if comparison to the the past or peers is possible), goals cannot be effective without measures. To guide progress toward a goal, goals must be backed up by a robust measurement system that tracks state (or local) progress toward the goal and allows the identification of effective and ineffective strategies. Robust measurement systems can drive progress even without goals if they enable agencies and others to assess the incidence of problems and objectively identify effective government interventions. Measurement systems can be especially effective when they compare performance, because some governments like to be leaders while pressure can be placed on those at the "back of the pack." With comparisons, pack leaders in effect serve as de facto goal setters.

Recommendation 3: Standardize and Normalize

Federal agencies should play a role in ensuring cross-state data are standardized and normalized—that is, information submitted from each state for a single data category has the same meaning and is characterized in units that enable appropriate comparisons. States or their associations should play a leadership role initiating standardization when federal agencies have not focused on the issue. Federal agencies can support that effort by staffing, supplying information to, or serving as the secretariat for state-led efforts to set standards. When neither the federal government nor the states carry out the needed normalization, non-governmental organizations can play a catalytic role, conducting analyses that can serve as a model for future federal or state efforts.

Recommendation 4: Require Measurement

Standardized state performance measurement is more likely to happen if Congress mandates its generation, collection, and dissemination. When common metrics do not already exist, Congress should require and financially support the full spectrum of state performance measurement, along with annual training of state information handlers, in areas where federal agencies depend on states to accomplish their objectives.

Recommendation 5: Involve and Benefit Those Being Measured

Federal agencies should routinely engage states as co-owners of performance data and as decision makers in developing tools to enhance the use of the data, especially to serve state needs. When federal agencies lack authority to set national standards, they should encourage states to develop their own data and performance standards and support states' efforts to do so, while maintaining sufficient influence to assure that state-set performance standards continually improve social outcomes relative to existing standards and conditions.

Recommendation 6: Encourage Analysis

The federal government should conduct its own analyses and encourage third-party analysis with federal funding. National associations of state officials can also play a leadership role organizing and managing co-investment in the development of analytic tools. Public interest groups can play a crucial role contributing to improved performance by conducting analyses that answer questions practitioners and the public want to know. Although some NGO analyses will inevitably spark volatile reactions by those being measured, the NGOs should take care to present their findings in a constructive way, commending good government performers, not just criticizing weak ones.

Recommendation 7: "Market" the Results

Federal agencies should make information dissemination in an audience-focused format a priority. They should make it a priority to return the information they gather to those who supply it (states, local governments, etc.) with value added, helping them learn from the experience of other states and even from their own. They should also share performance information with Congress in a format that quickly answers questions Congress has. This may imply displaying performance information with greater geographic specificity so those in Congress can relate to it. This also implies delivering information in a timely manner, especially when the information is needed to inform policy debates.

Recommendation 8: Motivate with Comparison and Rewards, but Carefully

Federal agencies should hone their skills to provide balanced presentations of comparative performance information, including developing the skills and capacity of their regional offices to coach and encourage friendly competition among small groups of states. In addition, Congress and federal agencies should explore increased use of incentives structured similar to those of the seat belt program, which mandate comparable measurement and reward performance relative to peers and past performance.

Recommendation 9: Share Best Practices

Finally, the federal government can serve as a valuable expert resource for state and local governments, aiding them in the search for effective practices. It can look for and conduct research that governments "on the ground, overwhelmed just by getting the job done" find hard to do. Regional and state offices of federal agencies may be especially well suited to this task, especially when reinforced by central office support scanning across the country and the world for more effective and efficient ways to improve societal outcomes.

Introduction

Federal agencies that set goals for or measure the performance of states often find themselves in testy territory. For both political and practical reasons, states resent efforts by the federal government to influence their goals and their performance levels. Nonetheless, citizens often turn to their elected representatives in Congress to require federal agencies to drive state performance improvements when they feel their state or local governments have failed to address a problem necessitating a governmental response.

The question is: How can federal agencies best use performance goals and measures to work with state and local governments to improve societal outcomes? What can they do that is both practical to implement and politically feasible? The federal government has long used measurements of conditions and practices in states, together with mandates and money, to encourage both the measurement and improvement of social results. A better appreciation of federal agencies' experience in this arena can help Congress craft better laws, help federal domestic policy agencies design more effective implementation strategies, and help state agencies reap greater value from their frequent interactions with the federal government.

Results Act Is Changing the Conversation

Now that all federal agencies are required by the Government Performance and Results Act of 1993 (GPRA) to set outcome-focused goals and report annually to Congress on progress toward those goals, the conversation about performance measures and goals is changing between states and the federal government. (See "How Government Organizations Benefit from the Use of Goals and Measures.")

Most federal departments and agencies charged with implementing domestic programs—especially the Department of Education, the Department of Transportation, the Department of Health and Human Services, the Department of Housing and Urban Development, the Environmental Protection Agency, the Department of Agriculture, and the Department of Labor—depend heavily on other levels of government to accomplish their goals. Nearly a quarter of federal domestic program funds are transferred to state and local governments through these seven departments and agencies.This funding figure, however, fails to capture the full magnitude of the relationship, because most federal agencies also influence the ways state and local governments use their own funds through grant terms, regulations, and other requirements. Under GPRA, federal domestic agencies will inevitably need to grapple with whether and how to use state and perhaps local government performance measurement. Surprisingly, GPRA says little about how federal agencies should integrate information about state and local performance into their GPRA reports. Given federal reliance on states and localities to accomplish their programmatic goals, GPRA's silence on the subject of states is surprising, as is the limited discussion of this subject in the policy guidance documents issued by the U.S. Office of Management and Budget pertaining to GPRA implementation.

How Government Organizations Benefit from the Use of Goals and Measures

Goals

Goals can strengthen democratic accountability and improve outcomes through motivation, information, and focus.

- The articulation of goals allows Congress and the public to know what an agency believes its priorities are in the context of all the possible work the agency could do to advance its mission and fulfill its legislative mandate. If Congress does not agree, it can realign agency priorities and congressional expectations through new authorization measures and limits on spending in the annual appropriations process.
- The motivating power of goals can improve societal outcomes. Goals tend to motivate because people respond positively to challenging but realistic goals. A small number of stretch goals can be even more motivating because they force those trying to reach the goals to rethink current implementation strategies and seek innovations to realize dramatic productivity gains. When goals are linked with rewards or sanctions, they can be even more motivating, provided the rewards or sanctions are not so high that they encourage perverse behavior, including distortion of the measurement system, and that sufficient attention is directed to assuring measurement honesty. Linking resources to goals can be a means for gaining necessary appropriations, or at least for the setting of realistic expectations for program achievement.
- Goals also help agencies focus internal energies and enlist assistance from those outside the agency whose actions directly affect outcomes. When individuals within an organization work toward disparate goals, they can easily miss opportunities for synergies and economies of scale. Goals focus agencies and help them reap synergies from internal collaboration. When agencies depend on others outside the organization to achieve success, the articulation of a specific goal for a specific place and for a specific kind of business (or group of people) by a specific date can be used to rally external assistance and cooperation. Essentially, goals play a communication and attention-focusing function that is especially valuable for larger organizations.

Measures

Government organizations can use performance measures, with goals and on their own, to strengthen democracy, motivate, and help organizations discover more effective and efficient ways to achieve improved societal outcomes.

- Reporting progress toward goals in annual performance reports communicates to Congress and the public how well an agency is achieving its goals. This provides an additional opportunity for citizen and congressional feedback. Congress can hold hearings about an agency's performance or communicate more directly to agency leaders questions about and ideas for improving performance. Citizens can bring attention to the agency's performance and their support for or concerns about it through letters to the agency or the newspaper, opinion pieces, press events, rallies, and various other means. They can also try to influence the agency's behavior through more direct contact or indirectly through the electoral process.
- Comparative performance measurement, even without goals, tends to motivate because most people have a competitive instinct and like to do well relative to their peers when they perceive comparisons to be fair. And, while only a few may strive to lead the pack, no one likes to trail at the back. Also, trend information, essentially comparison over time, stirs the intrinsic instinct to beat one's own personal best.
- Standardized performance measurement helps identify places and programs achieving stronger outcomes as well as those having trouble. Studying the strong performers, often dubbed "benchmarking," does not necessarily identify best practices, because stronger performance may be explained by conditions or characteristics not related to the actions of government. Studying the strongest performers can, however, lead to the identification of the factors that do explain strong performance levels, such as the different actions taken by particular regional, state, or district government offices. These successful practices can then be continued or replicated. Comparative measurement also leads to the identification of areas with weaker performance, triggering follow-up questions to determine if the program intervention being used does not work and should be terminated, or if those running the program need assistance from those experiencing greater success. In sum, comparative performance measurement provides a continual mechanism for finding problems that need fixing and intervention strategies worthy of continuation or replication.

There is strong and growing evidence that government agencies that successfully integrate performance goals and measures into their strategic and daily operations realize significant performance gains. What is also incontrovertible is that any organization that fails to measure progress toward organizational objectives cannot possibly manage to them, or even assess if it is making progress toward them at all.

Adapted from Harvard University Kennedy School of Government Executive Session on Public Sector Performance Management, Get Results Through Performance Management: An Open Memorandum to Government Executives (Visions of Governance for the 21st Century Program, Kennedy School of Government, Harvard, 2001), available at www.ksg.harvard.edu/visions/performance_management.

Key Questions Regarding Federal Use of State and Local Performance Measures

1. Should federal agencies adopt clearly defined outcome goals for states or require states to adopt them?
2. Should they require public reporting on progress toward those goals?
3. Should they require performance reporting that is comparable for all states?
4. Should the federal government enter into formal performance agreements with each state encompassing these goals and measures?
5. What should federal agencies do with the performance measures, once reported?
   a. Should they publicly report them?
   b. Should federal agencies compare state performance?
   c. Should they analyze them to find the successes worthy of replication and the problems warranting intervention or assistance?
   d. Should the level and nature of federal actions in a state, including rewards and penalties, be calibrated to state performance?
6. How should federal agencies handle goals that states have already set for themselves?
7. Should federal agencies include state-specific performance information in their annual GPRA reports?

The shift in government's attention to societal outcomes as an outgrowth of GPRA raises a new set of questions about the federal/state relationship. Should federal agencies change the way they currently work with states? Should they adopt clearly defined outcome goals for states or require states to adopt them? Should they require public reporting on state progress toward those goals? Should they require performance reporting that is comparable for all states? What should federal agencies do with the performance measures, once reported? How should federal agencies treat goals that states adopt for themselves? A few recent studies have begun to probe these questions.[1] Frederickson, for example, in a recent study of GPRA, urges federal agencies to "use GPRA as the vehicle through which federal agencies expand their monitoring activities to include the universe of third-party relationships."[2]

How Should Federal Agencies Respond?

At least for the moment, however, it has been left to each federal agency to sort out for itself if and how to change the way it works with states post-GPRA. To some extent, the lack of guidance makes sense. Each agency has its own laws, concerns, and organizational patterns that have evolved over time governing the way it works with states, including whether it sets goals for states and what information it must collect. At the same time, federal agencies face many common issues in dealing with the states. The agencies could benefit from each other's experience and might also benefit from some standardized practices. They could also benefit from a clearer sense of the potential benefits of performance goals and measures and the ways federal agencies can use measures in working with states to improve societal outcomes.

On a practical level, in answering these sorts of questions, the federal government must consider the same issues that face any large organization, public or private. What is best done centrally and what is best decentralized? Centralized actions often lead to economies of scale in research, production, and marketing. In many circumstances, centralized decision making can also trigger greater investment in innovative solutions than can decentralized purchasing decisions by multiple, decentralized buyers. For example, when the federal government sets a performance standard for all states, such as a maximum emissions limit for cars or highway construction performance standards, more businesses are likely to invest in new product development to meet those standards than if each state sets its own standard.

Counterbalancing the economies and innovation-inducing value of centralization are associated diseconomies of scale and innovation-diminishing costs. Henry Ford captured the essence of the costs of centralized decision making when he assured his customers they could have a Model-T in any color they wanted as long as it was black. National road quality standards may feel too stringent for sparsely traveled sections of Western state roads and too lax for heavily traveled roads in the East. National water quality standards appropriate for a river used heavily for recreational purposes in a densely pop-

ulated community may seem unreasonable for a stream that has water only two months a year and is located in a sparsely populated area. Centralized standards can also inhibit innovation when they make it more difficult for innovators with technologies that exceed national standards to find purchasers of their more innovative but not yet federally approved products.

The Political Dimensions of Measuring Intergovernmental Performance

Practical considerations pertaining to the federal use of state performance goals and measures tend to pale next to the political ones. Federal decisions about state goals and measures provoke fears that unbounded federal decision makers will impinge on constitutionally protected states' rights. The 10th Amendment to the Constitution reserves to the states or the people of the United States "powers not delegated to the United States by the Constitution, nor prohibited by it to the States." Neither the Constitution nor the 10th Amendment clearly articulates what those reserved powers are, however. From the beginning of U.S. history, an unending debate has raged about the path of the dividing line delineating the roles of federal, state, and local governments. Periodically, the Supreme Court weighs in to clarify the demarcation, but it often views that clarification as the prerogative of the U.S. Congress. This was the message the Supreme Court sent in *Garcia v. San Antonio Metropolitan Transit Authority (1985)* when it opined that democratically elected representatives, through the federal legislative process, should balance the debate regarding appropriate federal and state responsibilities.

It is often assumed that congressional balancing will follow party lines—that Republicans will favor state assumption of responsibilities with no federal constraints while Democrats will prefer centralized federal authority. In reality, the votes of federal legislators tend to align more with their support for the particular federal policy being debated rather than strict adherence to principles of states' rights. The Republican-led welfare reform law, for example, whose legislated purpose was to "increase the flexibility of the states in operating a program" to assist needy families is replete with requirements restricting state flexibility in determining whom states can help and how much support they can provide.[3] At the same time, Democrats fight mightily to assure that the federal government does not restrict state flexibility in legislating tort laws. In short, it has never been possible to sort out the balance between the constitutional partners in a rational way that would prevail under all circumstances.

Federalist Paper No. 51, written by James Madison, suggests the ambiguity may be intentional:

> In the compound republic of America, the power surrendered by the people is first divided between two distinct governments, and the portion allotted to each. Hence a double security arises to the rights of the people. The different governments *will control each other,* at the same time that each will be controlled by itself. *[italics added for emphasis.]*

Multiple levels of government with ambiguously defined spheres of authority each check the powers of the other, protecting citizens. At the same time, they also serve as competitors of a sort in what could otherwise be a monopoly situation, giving citizens the ability to press for specific products and services from one level of government when the other refuses to address a problem citizens believe warrants a governmental response.

The politics of the federal/state relationship is not only a matter of political philosophy, however. It is also affected by personal sentiments. When a federal official makes a decision forcing a state or local official to take an action he or she would otherwise not have taken, it naturally unleashes feelings of resentment by the person being compelled to take the action toward the person wielding the power, especially if handled undiplomatically. And, when federal and state decision makers are of different political parties or seek the same political office, federal goals and state measurement provide ammunition for political advancement. Federally sponsored measures that encourage citizens to judge the performance of their state agencies are seldom welcomed by state officials.

The Integrating Role of Federal Agencies

The federal government's success using performance goals and measures to work with states and local governments to improve social outcomes depends on agency skill in balancing the use of three tools it has at its disposal: measurements, mandates, and money. Selected experiences from several federal agencies suggest fruitful performance management practices that federal agencies can adopt to work more constructively with state and local governments to deliver improved societal results to the public.[4] They also demonstrate the inevitable tensions in and difficulties of an intergovernmental service delivery system.

Federal agencies can take a wide range of beneficial actions using goals and measures with states to improve outcomes. These include information collection; analysis to find successes, problems, and patterns of problems; audience-tailored organization, presentation, and dissemination of the raw data and analyses; broad and easily accessible dissemination; serving as an expert resource; goal setting linked with incentives for motivation; and fair peer comparisons for the same purpose. Congressional mandates to carry out these actions greatly enhance their prospects for success, although they are not necessary.

As a general rule, this study finds that federal agencies that emphasize the informational value of performance measurement—that add value to the information they gather by organizing and analyzing it to identify successful practices and detect problems or at least trigger focused follow-up questions—are likely to build more sustainable and valuable measurement systems than measurement systems used primarily to assure fulfillment of state and local commitments to the federal government. By their actions and the way they use performance information, federal agencies can encourage the creation and maintenance of a learning environment that continually seeds, harvests, and re-sows the lessons of state experience. This is not to suggest that federal agencies should not use performance goals and measures to boost accountability; rather that the emphasis in using goals and measures should be on building an informational system that helps all governments more readily learn, from their own and each other's experience, ways to improve social outcomes. Louis Brandeis is oft-quoted proclaiming the states as the "laboratories of democracy." Laboratories produce little of value, however, without someone in the laboratory objectively documenting, analyzing, and writing up the experimental results.

The following case studies of four federal agencies that depend heavily on states and localities to meet their GPRA goals provide examples of how federal agencies have attempted to use—to varying degrees of success—measurement, mandates, and money to pursue national goals in partnership with states and localities.

Case Study One: Environmental Protection Agency—A Focus on Outcomes Via Partnership

Federal environmental laws have long embraced the idea of setting environmental goals and measuring progress toward them as a way to improve environmental conditions. The Clean Air Act, for example, requires states to meet specific air quality goals by specific dates and sanctions states with heavy penalties if they fail to meet them. The Clean Water Act declares that "it is the national goal that the discharge of pollutants into the navigable waters [of the United States] be eliminated by 1985" and further that "wherever attainable, an interim goal of water quality which provides for the protection and propagation of fish, shellfish, and wildlife and provides for recreation in and on the water be achieved by July 1, 1983…." In addition, both laws require annual reporting on air and water quality in each state.

Shifting the Federal/State Relationship

Despite the clear establishment of environmental goals and requirements to measure progress toward them in some of the major federal environmental laws, the Environmental Protection Agency (EPA) and state environmental agencies found it necessary, in May 1995, to adopt a policy to do what many outside the government assumed they already did—focus the state/EPA relationship on improving environmental conditions.[5] This policy, the National Environmental Performance Partnership System (NEPPS), aimed to shift the federal/state working partnership from one that had, over time, become more attentive to assuring the completion of specified activities than to improving the quality of the environment.

The effort to adopt and implement this policy provides an instructive place to begin our exploration because it imparts a sense of the practical and political issues that arise as federal agencies and states grapple with federalism, goals, and performance measurement. These implementation challenges are not unique to EPA. Most federal agencies and their counterpart state organizations face problems similar to those that led to the adoption of NEPPS. The way EPA and states choose to tackle those problems has been considered a model for others because the states were so intimately involved in crafting the proposed solution to the problems, performance partnerships. Despite the joint design of the partnership system and progress putting it into practice, however, numerous implementation issues have arisen that remain to be resolved.

Reason for NEPPS Adoption

NEPPS evolved in response to state frustration with the way EPA conducted oversight after it "delegated"[6] authority to a state to implement federal law. Under most federal environmental laws, EPA delegates to a state the authority to implement federal law when a state meets minimum specified conditions. These conditions generally include adequate organizational capacity, state laws and regulations at least as protective as the federal ones, adequate financial resources, and appropriately skilled personnel.[7] When a new law passes, states petition to assume delegation responsibilities for the whole law or a subsection of it, and EPA assesses whether a state meets the conditions for assumption and decides whether or not to award delegation authority. Where a state does not seek

delegation or EPA does not award it, EPA retains authority for direct implementation of that aspect of federal environmental law in the state.

After authorizing a state to implement federal law, EPA maintains oversight responsibility.[8] EPA's oversight practices have evolved over time to include the review of selected individual cases (e.g., whether a single permit or enforcement action was properly handled), confirmation that agencies delivered activities promised in annual grant agreements (e.g., the number of inspections conducted), and assurance of fiscal accountability (e.g., that funds were spent honestly and in accord with federal spending requirements.) Although EPA's oversight activities have varied over time and across regions, for the most part, prior to NEPPS adoption, the agency paid little attention to changes in the level of environmental harms (e.g., discharge or emissions levels, toxics released to the environment), improvements in environmental (e.g., ambient air and water) quality, or compliance levels of different programs in each state, even though it collected data relevant to these concerns.

EPA's management of the air quality program was a noteworthy exception to the case and workload orientation of the oversight process. Federal law requires every state to meet minimum national air quality standards for six pollutants; the law requires EPA to penalize states that fail to meet the standards by holding back federal highway moneys. Other noteworthy exceptions exist, as well, such as the effort to improve water quality in the Chesapeake Bay. There, several states have successfully joined together and with the federal government to set environmental goals, measure and publicly report progress toward them, and improve water quality. The Chesapeake Bay regional effort operates outside and in addition to the traditional EPA/state relationship, however, although it arguably could be a model for it.

As states built the capacity of their own environmental protection agencies, they found EPA's oversight constraining. State leaders with strong visions about how they wanted to run their own programs experienced several different types of problems: clashing priorities, interference from their own managers, systemic barriers to innovation, second-guessing, a lack of attention to outcomes, the limited value of EPA compilations of state-supplied data, and inadequate use of available information. At the same time, states shared a concern with both environmental activists and the regulated community that EPA's approach to oversight was inconsistent across the regions and—a special concern of environmental activists—sometimes non-existent.

Clashing Priorities

EPA's effort to set annual priorities for the use of its funds often interfered with states' efforts to pursue their own priorities. For each of its 16 categorical grants to states, EPA program offices would prepare an annual guidance setting forth federal expectations for state use of the grants. This guidance would be incorporated into the annual workplan that regions negotiated with each state, setting forth the expected number of permit reviews and inspections, the types of facilities to be inspected to align with EPA's inspection priorities for the year, and reporting requirements. EPA's expectations often drove out state planning efforts, especially in states where EPA funds composed a significant portion of the state environmental agency's budget. If a state felt other needs were more pressing than those selected by EPA or recognized a time-sensitive opportunity afforded by local events and sentiments, it could not pursue that need or opportunity using EPA grant moneys without extensive negotiations with EPA to shift from the EPA priority to the state-preferred use.

Leadership Constrained by Fragmentation

Some of the problems state leaders faced resided in their own operations because minor fiefdoms had grown up around the individual streams of EPA money. Annual negotiations for each individual grant routinely took place between mid-level managers from both the state and EPA. When state leaders wanted to combine grants to pursue a single project, say a pollution prevention effort targeted at a single industry, state staff could resist a shift of funds by citing the conditions of the EPA grant.

Financial Accountability Systems Impeded Innovation

EPA's need to link activities to specific fund accounts to assure fiscal accountability also slowed state innovation efforts. Massachusetts' experience

with one project illustrates the problem. Massachusetts wanted to change its inspection practices to make them more effective and efficient. To assure compliance with permit conditions, it wanted to conduct a single inspection for facilities holding multiple permits (for air, water, and waste) rather than multiple inspections for each permit. A test of this approach by the state showed that it would save inspection costs both for the state and for the regulated party without compromising the value of the inspection. Also, it would prevent another problem that sometime arose: permit inspectors, unaware of the permit conditions of other programs, recommending solutions for one program that created problems in another. EPA had long advocated this sort of "multi-media" approach in policy. Nonetheless, it took Massachusetts more than three years to win EPA's approval for multi-media inspections.[9] A key sticking point was sorting out how the state would charge its multi-media inspection activities back to individual program office grants, each of which had its own separate grant conditions. Despite evidence from the pilot that the program would be more environmentally effective, the system established to assure the accountability of federal grant funds erected barriers to innovation. Only the most tenacious states kept trying to innovate, instead of simply agreeing to use federal money in line with EPA expectations.

Second-Guessing

States also disliked when EPA initiated an enforcement action in the state without prior notice to the state or when the EPA action repeated, undid, or otherwise interfered with work the state was already doing with a regulated party. The state of Maryland encountered this sort of problem in its efforts to improve drinking water quality in several small communities. For several years, Maryland had been working with a few small towns that were out of compliance with federal drinking water standards. The residents of the towns had very low incomes, and the state sought a realistic solution that would fix the public health problems the environmental laws were designed to address at a price the low-income towns could afford. Despite significant state progress working with the community to devise a solution that protected the health of community residents, the EPA regional office initiated its own enforcement case. This infuriated state program managers who had devoted extensive time and effort to building trust with the community to get it to deal with its problem. EPA eventually agreed to support the state's approach, but not without significant state time devoted to convincing EPA to drop its enforcement action. While a mid-1990s survey conducted by the states revealed that EPA "over-filings" were relatively rare, they tended to enrage affected states when they occurred.

Environmental, Health, Compliance, and Capacity Information Absent from Oversight Review

EPA seldom considered changes in ambient conditions, environmental impacts (emissions and discharge levels), or compliance rates as the basis for its annual state grant negotiations and oversight reviews. Except for ambient air conditions related to the six criteria air pollutants, EPA did not even routinely analyze the environmental conditions or impact information it collected from the states. Nor did most regions routinely monitor changes in state laws and resource levels to assure each state continued to satisfy the terms of its delegation agreement.[10] EPA tended to wait until citizens groups petitioned the agency to withdraw delegation, after which the agency would initiate a review of program adequacy.

Instead, because federal funds depended upon them, priorities set forth in the annual grant guidances issued by individual EPA program offices and the annual workplans served as the basis of annual reviews. A state might theoretically achieve significant reductions in discharge or emissions levels or improvements in ambient conditions but still get lambasted by EPA (especially the inspector general) for failure to carry out activities to which it committed in its annual grant agreement.

Garbage Out, Garbage In

One of the most frustrating aspects of the state/EPA relationship concerned the way EPA handled the data that states reported to it. States have long reported extensive amounts of information to EPA pertaining to inspection levels, compliance status, ambient air quality (some of which EPA collects itself), water quality, discharges from wastewater permit holders, and air emissions. EPA compiles these data for the specific reports it is required to

submit to Congress, most notably an annual air quality report and a biennial water quality report. It does not, however, compile state data into a single report showing the full scope of what the agency knows about environmental or compliance levels in the state. Much valuable data that could be used to generate environmental impact and compliance trends for all states or for groups of facilities within a state reside within EPA databases, with only limited after-the-fact analysis to glean the lessons of the information.

EPA has occasionally conducted analysis of state-specific performance data for its own use—to determine whether to grant a state request for program delegation, respond to an interest group petition to withdraw a state's delegation, or decide program priorities for the coming year. EPA has seldom shared these analyses with states and seldom made them public, fearing negative state reaction to EPA reports that might be considered evaluative. EPA's fears grew from years of experience. Not every state welcomes EPA analysis and the resultant pressures it might put on the state, and a few often protest vociferously when the analyses are released.

With little data coming back to states from EPA, many states built their own information systems over the years so they could more easily retrieve information they needed to manage their programs. As states built their own systems, system incompatibilities arose, requiring many states to enter data separately into both EPA's systems and their own. Not surprisingly, since states primarily relied upon their own systems, they spent little time worrying about the quality of the information in EPA's systems. A spate of analyses on state compliance and inspection programs conducted by public interest groups in 1999 and 2000 using EPA's compliance databases made clear how seriously flawed data in some of EPA's systems were.[11]

Inconsistent or Non-Existent Oversight

EPA's oversight practices were highly inconsistent from region to region. Not surprisingly, this was hugely irritating to states in regions with more aggressive oversight practices. Because of this variation, many states perceived EPA oversight more as a reflection of individual personalities capriciously threatening to exercise their power to withhold federal funds than as consistent national policies designed to improve environmental quality and fairness. States were not the only ones who saw problems in the way the system operated. The regulated community and environmental activists were also concerned about the inconsistency of EPA's approach to oversight; activists often felt EPA oversight was inadequate.

In sum, after states began to build the capacity of their own environmental protection organizations and develop their own strategic priorities, the traditional EPA approach to grant giving and oversight began to chafe. State leaders wanted change. They wanted a system that gave them more flexibility to set their own priorities, and more credit for what they had accomplished. And they wanted a system that focused more on the environmental and non-compliance problems in each state, rather than on the activities that more than a dozen individual EPA grant-giving programs selected as national priority activities each year.

Using Performance Goals and Measures in Working with States

With NEPPS, states and EPA jointly tried to answer many key questions regarding federal use of performance goals and measures in working with states.

- *Environmental goals and measures.* NEPPS calls for EPA and states to reach agreement on a common set of performance measures every state would report for purposes of national environmental assessments. In addition, each state that opted to participate in NEPPS would identify appropriate state-specific environmental performance goals and measures. Both the common and state-specific measures would inform decisions about state and EPA activities in each state each year. The states would work with EPA in an equal partnership to select, test, develop, adopt, and update the measures.
- *State self-assessments.* Under NEPPS, states are expected to conduct self-assessments and share them with the public. These would include an assessment of key environmental problems and opportunities in the state; a description of the recent performance of state programs including

an analysis of program weaknesses; an assessment of fiscal accountability; the state's proposed action plan for maintaining and improving environmental conditions in the state; suggestions for ways EPA could assist the state to improve performance; and a report on the state's effectiveness implementing the prior year's environmental performance agreement.

- *Performance agreements.* Based on both the state's and EPA's assessment of environmental conditions and state program performance, each state and EPA would sign an agreement regarding appropriate national and state-specific environmental goals, program performance indicators, state commitments for specific deliverables and activities to address identified needs, disinvestments, and federal commitments. The agreement would take precedence over the traditional program work planning process.
- *Differential oversight.* EPA would shift the focus of its individual state evaluation from the review of and intervention in individual permits, inspections, and enforcement actions to after-the-fact, program-wide audits and assessments. Based on those reviews, EPA would exercise differential oversight. That is, it would treat state programs in different ways, calibrated to the strength of state program performance. EPA would return to case-specific review and intervention for state programs demonstrating poor performance, but continue conducting only after-the-fact reviews for programs with strong performance. Exceptions might be made when problems cut across multiple states or when local politics necessitated action by a government entity other than the state.

  As a component of differential oversight, the NEPPS system called for the creation of an official designation for state programs that consistently demonstrated strong performance. Such programs would be designated "Performance Leaders" if states requested the designation based on criteria EPA and the states would jointly define.
- *Public outreach and involvement.* EPA and the states committed to disseminating to the public information about the NEPPS system, the state self-assessments, and the annual state/EPA performance agreements. Public comment on NEPPS and its implementation would be invited.
- *System and state evaluation.* EPA and the states agreed to continual review and improvement of NEPPS to assure effectiveness, public credibility, and fiscal soundness. In addition, the NEPPS agreement called for the states and EPA to explore the use of visiting program evaluation teams composed of both state and regional staff to review each state's programs.
- *Performance Partnership Grants.* Concurrent with and in support of NEPPS, EPA sought and obtained congressional authorization giving each state the option to combine federal grant money from 16 categorical EPA grants into a single grant, essentially creating a state-triggered, discretionary block grant.

Progress After Eight Years

Since the adoption of NEPPS in 1995, both states and EPA have made noteworthy progress addressing the problems the NEPPS system was intended to address. The progress is evident in two forms: what states have been able to do individually because of NEPPS and what changes have been made to implement the NEPPS framework.

NEPPS has provided state leaders who want to pursue environmental priorities other than those EPA has selected with a ready mechanism for making that possible. In the early years of NEPPS, for example, Florida conducted a self-assessment showing that by following EPA's guidance for inspections, the state devoted significant resources to the inspection of facilities with historically strong compliance rates while paying insufficient attention to others posing a far greater environmental threat. Through the NEPPS process, the state was able to negotiate a more sensible deployment of resources. Numerous states, including Connecticut, New Hampshire, Illinois, New Jersey, Oregon, and Maryland, have successfully begun to shift agency resources to address serious local environmental and compliance needs that would not have been addressed simply by following EPA's program priorities or the state's prior year activities. Several states, including Oklahoma, Ohio, Illinois, Maryland, Florida, and the states in the northwest

region of the United States, have successfully negotiated reductions in reporting requirements.

Further, because the NEPPS agreements are posted on the Internet, the public can find far more information about state plans, programs, and progress than was previously possible. Although some of these plans are still quite bureaucratic in their language, they all share with the public information that had previously been hard to uncover about environmental conditions in each state, the programs and projects designed to address the conditions, and annual progress.

In addition to the progress of individual states, numerous changes have also been made to implement the NEPPS framework across the country.

Performance Goals and Measures

Each of EPA's five main program offices has successfully negotiated Core Performance Measures (CPMs) with the states to support NEPPS. In some cases, CPMs called for new data reporting by the states; in many cases, states and EPA simply reached agreement on the items states already reported that would be used for performance assessments.

While negotiating the CPMs, states expressed a strong concern that NEPPS would increase their "reporting burden" to EPA. After negotiating the CPMs, states asked EPA to work on reducing the reporting burden. States undertaking sophisticated upgrades of their own data systems were especially fearful that EPA would require them to continue reporting into the more antiquated EPA systems, necessitating at least double the data-handling workload.

To address that problem, EPA and states began to formulate an alternate vision where EPA would be able to sweep modernized state systems for the data EPA needed. To make that sort of data sweep feasible, states would all need to use common data definitions in their own information systems. In November 1999, EPA and the states reached an agreement to establish the Environmental Data Standards Council (EDSC). The Council, made up of EPA, the states, and tribes, "identifies those areas of information for which having standards will render the most value in achieving environmental results, prioritizes the areas, and pursues the development of data standards."[12] As of November 2002, the EDSC had approved 10 sets of data standards. In July 2000, the states and EPA committed to developing a blueprint for the new Environmental Information Exchange Network, the first version of which was released in October 2000.[13]

Performance Agreements

Two-thirds of the states have negotiated environmental Performance Partnership Agreements (PPA) with EPA; one-third have opted to stick with the traditional approach. A few states opted not to participate in NEPPS for political reasons. Others chose to delay their participation until they could assess the value of the system to the initial adopters. A few tried it and then decided to go back to the traditional approach.

The content and organization of the agreements varies dramatically by state. Some agreements present a thorough overview of all state programs and priorities, while others serve primarily as a statement of the operating principles for all the other agreements between EPA and the state.

Self-Assessments

States that sign PPAs are also expected to conduct self-assessments. Some states, including Illinois and Florida when it first implemented the performance partnership system, conducted thorough self-assessments that provided the public with an unprecedented amount of information on the state of the environment and agency management. Florida's report won finalist status in the Ford Foundation/Kennedy School Innovations Award program. Some states that had already done strategic plans used those as their self-assessments. EPA's New England regional office developed guidelines for self-assessments for its states that encompass the original language of the May 1995 agreement, calling for states to describe their environmental quality, program effectiveness, program priorities, and proposed strategies in their assessments. Some states have not taken as comprehensive an approach to the self-assessment, and EPA has not yet tried to assure that all regions and all states include all the components of a self-assessment identified in the May 1995 NEPPS agreement.

Information to the Public

The quality and quantity of information states provide the public has improved markedly since 1995. Far more states are now reporting to the public on state environmental conditions. This may be attributable to NEPPS, to the concurrent movement by states to adopt results-focused management, or to technology advances that have dramatically reduced the cost of disseminating information broadly.[14] The best NEPPS agreements or self-assessments give the public a much better sense than was previously available of environmental conditions in the state, state plans for improving the conditions, and the effectiveness of state efforts as does Illinois' FY2001 self-assessment.[15] Maryland produces an agency-wide "Managing for Results" report for the public documenting the state's progress toward numerous goals. It also produces an annual compliance and enforcement report. Connecticut, too, produces an annual compliance and enforcement report. The report begins with a description of selected environmental problems in the state and then describes how the state is using compliance and enforcement activities to address the problems. This report exemplifies how some state leaders are trying to shift their attention from activities to outcomes. New Jersey produces an environmental report card in tandem with its NEPPS documents, and has recently begun to generate reports showing compliance rates for all its programs.

Many NEPPS agreements and self-assessments continue to focus on state activities rather than environmental impacts, but even they make it much easier for the public to learn about state priorities and activities than was previously possible.

Revised Administrative Regulations and Policies

EPA has revised its regulations for all state grants to support performance-focused management. The new regulations apply both to Performance Partnership Grants (PPGs) created in concert with NEPPS and to more traditional, categorical grants. PPGs allow states to combine multiple categorical grants into a single grant, using authority Congress approved in support of the EPA performance partnership system that essentially functions as a discretionary block-granting mechanism for each state. Several states have used this authority to get rapid approval for combining grant funds to run resource-saving, cross-media programs similar to the one Massachusetts took three years to negotiate.

Just as important, the Inspector General's (IG) office has supported the implementation of NEPPS by shifting its attention from regional and state compliance with activity requirements to regional and state use of outcome-focused information to guide their actions. This represents a dramatic shift and important signal from the EPA IG in support of increased use of outcome-focused goals and measures, especially since many feared the IG would erect significant barriers to NEPPS implementation.

Challenges

At the same time, many challenges remain. In coming years, EPA will need to work with states to develop better ways to organize and analyze core performance measures and other state information reported to EPA to make it more useful to EPA, the states, and the public. Without that information, it will be difficult to identify, in a fair and consistent manner, states with poor performance warranting more aggressive federal intervention and those with whom EPA should work more in partnership. It is also difficult to identify effective state programs worthy of replication. EPA will also need to encourage states to carry out more complete and consistent self-assessments to inform the public. Further, more state leaders will need to take up the challenge of guiding their own agencies toward a management emphasis focused on improving environmental conditions and raising compliance levels.

Summary

In sum, EPA and the states adopted the National Environmental Performance Partnership System to heighten organizational attention to environmental conditions in both federal and state agency decision making. While several federal environmental statutes establish environmental goals and require each state to measure some aspects of environmental and compliance conditions in the state, EPA and state management of these laws had evolved in a manner that emphasized processes over environmental performance.

EPA and state environmental agency leaders adopted NEPPS as a way to make clear that the

federal government should use outcome indicators as the dominant criteria for assessing state environmental performance. Further, state leaders were encouraged to shift the organizational focus of their own organizations toward setting environmental goals and managing to meet them. In addition, NEPPS signaled that the EPA/state relationship should evolve more into a partnership dedicated to improving environmental conditions rather than be an oversight process focused primarily on complying with individual statutory process requirements.

Since the adoption of the NEPPS agreement, states and EPA have made significant progress. They:

- Jointly adopted a policy encouraging interested states to set their own performance goals.
- Agreed on the core performance measures for assessing state performance.
- Established a new data standards council to set and update data standards on a regular basis.
- Convinced Congress to support a change to federal grant law that makes it possible for states to combine federal grants to deal with problems as the states face them.
- Revised federal grant regulations to support states that want to set their own performance goals and manage to them.
- Greatly increased public access to information about state environmental performance and plans to improve it.

In addition, a number of states have made great strides in basing more of their management decisions on information about environmental and compliance outcomes, often upgrading their information systems to support that effort. Also, several non-governmental organizations (NGOs) have issued reports providing thought-provoking models that suggest ways EPA might routinely organize and analyze state data in the future.

Case Study Two: The Federal Highway Administration—A Value-Adding Information Operation

Since the first federal road office was established in 1893, the federal government has made performance information the center of its strategy for working with states on transportation-related programs. From its inception, the federal roads office gathered information from local and state governments about the characteristics of rural roads.[16] In 1904, the federal Office of Public Roads Inquiry conducted the first national inventory of all rural roads in the United States, gathering data about road mileage, as well as information about taxation, sources of revenue, road laws, and expenditure levels in every county.[17]

The federal roads office long emphasized not only the gathering of information but also analysis and dissemination of that analysis to potential users, most notably the states. Soon after it began gathering information about road characteristics, it supplemented it with experiments in the laboratory and the field about the performance of different kinds of roads and bridges. It also placed a priority on broadly and quickly disseminating what it learned. Within its first two years of operation, it published 18 bulletins and 23 circulars. In May 1918, the federal Office of Public Roads and Rural Engineering launched the magazine *Public Roads,* with this purpose:

> It will be our earnest effort—always with the support and cooperation of the highway organizations of the States—to present matters of special interest to those directly concerned with the construction and maintenance of roads, to bring to all the progress of road improvement throughout the country, [and] to discuss its problems and record its results. Always with the single purpose and devout hope that from this closer association will be born a determined and united disposition to bring to road betterment that which is best in and for this generation, that which, in this period in our history, will make for the greatest strength of our Nation.[18]

This early emphasis on gathering, analyzing, and disseminating outcome-focused, user-oriented information created a cooperative, information-based culture that still characterizes the way the current Federal Highway Administration (FHWA) works with states. (See Appendix III for a more complete discussion of the historic approach to state performance measurement taken by the federal roads office, which built the foundation for the agency's current approach.)

Today, FHWA continues to publish numerous highly regarded and broadly used reports that help other government agencies improve the conditions and functioning of their roads. Building on the nationwide road surveys begun in the first half of the 20th century, FHWA began publishing in 1945 an annual compendium—*Highway Statistics*—providing detailed information for each state about the ownership and use of motor vehicles; receipts, expenditures, and road funding mechanisms; and the extent, characteristics, and performance of public highways and local roads.[19] Since the late 1960s, FHWA has also published a biennial report on the conditions and performance of the nation's highway and transit systems, showing national

trends and needs.[20] The agency continues to publish *Public Roads,* and recently recommitted to the mission articulated for the magazine when it was launched over 80 years ago of working closely with states to advance the shared federal and state objective of road improvement. In 2000, FHWA initiated an online version of *Public Roads* to bolster its ability to share relevant knowledge with interested users.[21]

Voluntary, Value-Adding Use of Performance Measurement

Congress requires the Secretary of Transportation to report on the conditions and performance of the nation's roads, but does not explicitly mandate the compilation of the annual highway statistics report with state-specific data nor the specific kinds of information state or local governments must report to the federal government pertaining to road conditions. FHWA establishes state reporting requirements through regulation. Despite the absence of federal law mandating state reporting or a link between federal funding and the submission of measurements, states have willingly delivered the data to the federal government for over 50 years. One possible explanation for state cooperation in data delivery is that by gathering and organizing data from all 50 states, the federal government returns to the states the information it collects in a format more valuable than when originally submitted. One current FHWA official describes how the state reporting relationship works:

> The fact is that it is somewhat voluntary. We have a long relationship with the states and, as a rule, they provide what we ask. In truth, when states refuse, we don't have much power to require that they submit the information. We have had a few instances where a state refused to provide information for a while. One state, for example, got irritated when a congressman complained that our reports showed poor pavement conditions in his district. At first, the state didn't want to give us any more information. Pretty soon, though, it started to give us data again.

Despite the lack of compulsory reporting power, states have tended to cooperate because the information they deliver to the federal government gets returned to them with amplified value, as noted earlier. The FHWA official elaborates:

> We have been doing highway statistical information for 50 to 60 years, and have turned it back from beginning. We have always done some value-added work when we turned it back to the states. It lets them see what other states were doing, and they see the data as a really valuable resource.

When problems erupt, the states tend not to direct their fury at FHWA, because they so highly value the information and trust that FHWA will handle it fairly, even if others do not. Again, an FHWA official observes:

> Sometimes, we get caught in the middle when the press uses it, although in almost all the cases, the analysis was done by someone else. Even there, our conversations with state folks are not antagonistic.

Beginning in 1984, for example, a University of North Carolina (UNC) researcher began releasing reports using FHWA-generated data to compare the efficiency of state programs.[22] The report infuriated the states because it compared states with very different characteristics without taking those differences into account, and then ranked all 50 states against one another. States with heavier truck traffic, harsher weather, or more bridges would inevitably face higher costs per road mile than those with less severe challenges; the UNC analysis failed to recognize these differences. FHWA initially issued a public response critiquing the report and supporting state concerns about the analysis. It does not, however, try to respond to every analysis of federal highway information (nor to subsequent releases of the UNC analysis) because it lacks the resources required to respond to every analysis and does not wish to give undue prominence to weak analyses.

To facilitate more accurate comparisons, FHWA now includes in the annual *Highway Statistics* report a section entitled "Selected Measures for Identifying Peer States." These data make it easier for states and other analysts to benchmark states more accurately with each other.[23] FHWA also

posts its guidelines for submitting data in each category used in the *Highway Statistics* report online, making it easier for the relevant state employees (and the public) to find the standardized definitions.

The way FHWA compiles *Highway Statistics* may minimize the prospects for simplistic performance comparisons, as well. The report contains reams of state data presented in numerous tables. The sheer volume of the data arguably has a de-politicizing effect because it makes obvious how many dimensions of variation may account for state performance differences. Moreover, FHWA makes no effort to conduct its own performance assessments of the states. In fact, the *Highway Statistics* report steers away from performance issues such as congestion and avoids ranking states in its tables. FHWA now produces an annual report presenting not only performance trends but also progress toward performance targets, but it focuses on national performance, not that of individual states.[24]

Federal Funding of Audience-Focused Analysis

FHWA has also been working with states for years to build computerized management systems that harvest the contents of the performance measurement databases it has created to make them more useful to states even as they serve federal planning purposes. Pavement management systems, dating back to the 1960s, help states evaluate alternative investment strategies for specific projects, rank projects for funding based on road conditions, schedule preventive maintenance work, and determine project replacement requirements. In the late 1970s and early 1980s, with funding from FHWA, Texas contracted with a vendor to develop a more sophisticated road management system to help it manage its roads, dubbed the Roadway Design System software. In 1991, FHWA funded a demonstration project for a bridge management system to serve the states, involving both states and private consultants in the design of the management system.

Federal Mandates Stimulate Private Investment

Federal mandates for state or local submission of comparable data and for the use of management systems can also stimulate private vendors to create audience-focused, value-adding computerized performance management systems. A private vendor developed Bridge Analysis and Rating System (BARS) software to enable states to respond to a federal regulation requiring each state to rate the condition of its bridges. Private sector consultants have also mined the federal database on local transit systems to develop and sell management systems that help metropolitan areas address their transit planning needs.[25] In these instances, federal mandates for measurement and use of the measurement created enough market demand to stimulate a competitive private sector response.

Cooperative State Investment in Performance Management

Even when federal mandates fail to stimulate the development of private sector products, federal seed funding can stimulate state co-investment in analytic tools that enhance the value of available performance information. The road and bridge management systems are particularly interesting; they demonstrate how federal action catalyzed cooperative state investment and ownership of systems to analyze data the states gather to inform management decision making.

Although Texas was very pleased with the road management software program it had developed with federal funding assistance, neither Texas nor FHWA felt it could support the ongoing maintenance and enhancement of the program. Texas decided to seek co-investors. With the American Association of State Highway and Transportation Officials (AASHTO), Texas approached other states to contribute to the upgrade and maintenance of the Roadway Design System. They initially raised $300,000, sufficient for the acquisition and upgrade of the software. About the same time, the vendor of BARS decided to discontinue its investment in the bridge product. Several state customers suggested to AASHTO that it acquire the rights to own, upgrade, and market the product.[26] With its ownership of the Roadway Design System and the Bridge Analysis and Rating System software, AASHTO launched AASHTOWare, the software arm of AASHTO. Today, AASHTOWare owns and markets seven separate design and road management systems.[27]

Mandates Rankle…

Despite the benefits to states of comparable performance measurement and its analysis, an experience in the early 1990s when Congress required state use of transportation performance data in management systems illustrates how much federal dictates can rankle. At the same time, it also demonstrates the staying power of useful performance measurement and management systems.

The Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA) required states to implement six separate management systems to improve the efficiency and effectiveness of federally assisted transportation activities. Transportation management systems, as defined by the General Accounting Office (GAO), are "tools that provide information to assist state and local decision makers in selecting cost-effective policies, programs, and projects to protect and improve the nation's transportation infrastructure." Management systems take a variety of forms, including computerized inventories of assets, software programs, systematic procedures or processes for collecting and analyzing information, and committees that develop recommendations to improve the systems' performance.[28]

ISTEA was not the first time the federal government had mandated state use of management systems. FHWA established a rule requiring every state to have a pavement management system for its major roads in 1989. It first mandated state adoption of bridge inventory systems, a precursor of bridge management systems, in 1968. ISTEA required states to adopt management systems in six areas: highway pavement, bridges, highway safety, traffic congestion, public transportation facilities and equipment, and intermodal transportation facilities and systems. ISTEA did not mandate new measurements. Rather, it expected states to adopt federally approved systems for using measurements to guide complex federal and state decision making. The federal government could withhold 10 percent of federal ISTEA moneys of the states that refused to adopt the mandated management systems.

Despite a legacy of the value-adding federal use of state performance measurement and the acceptance of several earlier federal mandates for transportation management systems, states balked at the expanded list of mandated management systems. At a time of heightened anti-Washington activism in state houses and city halls, states seized the opportunity to reject federal efforts regulating whether and how they incorporated data into their policy and priority-setting decisions. Four years after mandating state use of the management systems, a newly elected Congress repealed all but the mandate for congestion management systems in large urbanized areas.[29] In the opinion of one senior Department of Transportation (DOT) official, "States had less concern about mandatory reporting, but they got uncomfortable when we moved into mandates for planning."

… but Catalyze Constructive Change

Interestingly, despite their repeal, most states opted to continue developing their management systems anyway. In a 1996 review, GAO found that half the states decided to implement all six previously mandated management systems. All but two planned to implement at least four of the six systems; the remaining two planned to implement at least three.[30] The persistence of the management systems appears to be attributable to the simple fact that states find them useful.

GAO's investigation also revealed how the federal mandates provided an important impetus for change in some states. The federal mandate functioned as an external force acting on organizational inertia, waking a "body at rest" and putting it into motion to take advantage of an opportunity lacking a strong internal champion. For example, in response to the original ISTEA mandate, Montana decided to improve the timeliness of the data in its systems. Montana had previously been aware of the problems with using dated information for its own decision making, but lacked an impetus to fix it. The federal mandate provided that impetus. At the same time, when Congress eliminated the mandate, Montana slowed the timing of its upgrade. Similarly, DOT's interim rule calling for data integration across the six management systems pushed several states to create systems integration committees, bringing together relevant parties and data for the first time. New York's data integration effort quickly paid off: New York discovered that by resurfacing pavement with high-friction asphalt, it could reduce its accident rate by half.[31]

State Role in Standard Setting[32]

The cooperative approach FHWA has used to encourage the development of systems that enable states to reap greater value from the measures they collect is also evident in FHWA's work with states to develop performance standards. Building on a pattern established in the early 20th century by Thomas H. MacDonald, the federal roads agency director, FHWA engages states intensively in the development and review of technical design standards.

MacDonald, as head of a state road agency, had called on his fellow states to work together through AASHTO's predecessor organization, the American Association of State Highway Officials (AASHO), to develop common technical and performance standards. When he assumed the helm of the federal roads agency in 1919, MacDonald moved his idea into practice by calling on AASHO to form a permanent standards committee to help the federal government update its standards for roads, materials, and bridges under the new federal highway law. AASHO accepted.

Since few state legislatures were willing to fund state staff to work on projects that served other states, AASHO turned to senior federal officials to chair its technical committees. MacDonald himself chaired the AASHO technical standards committee its first three years. The federal agency provided the secretariat staff for the AASHO committees, as well. In addition, federal staff provided the background data used by the technical committees, including fleshed out proposals for discussion.

The precedent established by MacDonald is still reflected in current transportation standard-setting practices. AASHTO committees continue to propose their own technical standards, or guides, and federal officials continue to serve as committee secretaries in the highway area. AASHTO committees also comment on FHWA-proposed rules. Each state and FHWA decide whether or not to adopt the AASHTO standards in their own standards and subsequent designs and procurement requests. FHWA follows its own federal rule-setting procedures to determine whether to adopt the state-recommended standards as its own requirements for roadways using federal aid. These have tended to be in the areas of geometric design standards, bridge design standards, and traffic control device standards—and, more recently, with standards and architecture for electronic intelligent transportation system (ITS) applications.

Even with these changes, the structure of the relationship assures both parties strong influence and—when the federal government adopts standards similar to those recommended by AASHTO—reduces the states' ability to blame the federal government for one-size-fits-all federal standards.

Summary

In sum, by emphasizing the collection of state performance information, organizing the information for easy access by other states, analyzing it in ways that add value beyond what an individual state can learn studying its own experience, and supporting collective state ventures to enhance their own analyses of the information, FHWA has built an information-rich partnership with the states that continually drives performance improvements. Gathering and organizing the state performance information for easy public access also stimulates private investment and expertise to build knowledge and enhance outcomes. Emphasizing the value-added aspects does not eliminate the tension in the federal/state relationship altogether, but it seems to improve the prospects that the tension will be a constructive one.

Case Study Three: National Highway Traffic Safety Administration—The Challenge of Performance Mandates

As automobile usage increased, it created a serious new public problem: increased accidents and fatalities on the nation's roads. Again, Congress responded to the problem with an initial insistence on measurement. In 1966, Congress enacted the Highway Safety Act and mandated the establishment of an "effective record system of accidents … to determine the probable causes of accidents, injuries, and deaths."[33] In addition, the Highway Safety Act of 1966 required each state to establish a highway safety program in accordance with uniform standards promulgated by the federal government.[34]

Many credit Dr. William Hadden, the first director of the federal office charged with implementing the highway safety law and one of the congressional staff who helped craft it, with bringing a public health perspective to federal highway safety efforts. Just as public health professionals try to document incidents of disease, the conditions and behaviors that might explain the incidents, and the effectiveness of various intervention methods attempted, the National Highway Traffic Safety Administration (NHTSA) gathers data about traffic-related injuries and fatalities, the prevalence of behaviors and conditions that might explain why those events occurred, costs associated with the accidents, and government actions employed to improve traffic safety.

NHTSA makes information collection, analysis, and dissemination a key component of its work with the states, and is one of the largest data collection agencies in the federal government. Attention to the compilation and delivery of useful information has won NHTSA strong congressional support for its measurement efforts. One current NHTSA manager notes:

> Good, useful data generates support to produce more good, useful data. We can answer questions that people ask. Historically, our requests for budget increases have been supported because we are responsive and show that we can do the job.

Problem and Solution Identification

NHTSA concentrates much of its data-handling efforts on highway safety problem identification and on solution identification.

NHTSA uses after-the-fact incident analysis to understand the cause of the incidents. One study characterizes its problem identification work as "assembling data into sufficiently detailed subsets so that over representations of certain types of crashes or involvements can be isolated."[35] To obtain its data, NHTSA builds on accident data that state and local governments collect for their own purposes. It adds value to the information by amplifying that which is collected locally for accident follow-up, combining data across states, making it readily available for others to study, and analyzing it to understand the patterns of problems. NHTSA funds state employees in every state to review police crash reports, coroner's reports, registration data, and other relevant sources. The state employees code data about every fatality, creating a complete national database (Fatal Analysis Reporting System, or FARS) on highway fatalities.

To improve data quality, NHTSA provides and requires state employees to participate in a week of training each year.

Because of resource constraints, NHTSA tracks only fatality-linked data in each state, not accidents. NHTSA gathers information through national sampling for serious crashes (the National Automotive Sampling System, or NASS) and for all accidents (the General Estimate System, or GES). These sampling data are useful for understanding overall patterns of problems. They cannot be used, however, to detect patterns of high-incidence problems in individual states. All states maintain their own databases with information about all injuries, fatal and non-fatal. About a dozen of these states collect and code sufficient data about non-fatal crashes to be a useful complement to NHTSA's national accident databases.

Sometimes NHTSA sees a problem but cannot detect the likely causes from its own or state databases. In those instances, its routine analysis triggers more detailed follow-up inquiry. When, for example, one state showed a significantly higher rate of right-angle crashes than others, NHTSA looked more closely to understand the cause of the problem. It found that the state allowed more driveway curb cuts on its major thoroughfares than did other states. NHTSA is currently working with the Federal Motor Carrier Safety Administration to understand the cause of crashes involving trucks so appropriate safety strategies can be developed.

In addition to its efforts to gather data on problems, NHTSA also gathers and carefully analyzes data about government practices to address the problems. With data it collects about state program characteristics, program outputs, behavioral changes, and safety outcomes, NHTSA has acquired valuable knowledge about the effect of different government programs on fatality rates and injury levels. It knows, for example, that increased seat belt usage in automobiles, helmet wearing by motorcycle riders, and certain programs to discourage drunk driving can significantly reduce the number of fatal accidents.

To identify effective government interventions, NHTSA studies state-to-state variations in programs and performance. It can identify states that have the highest percentage of fatalities from drivers running off the road and those with a high rate of fatal accidents from right-angle crashes. Based on the evidence it gathers, NHTSA identifies the strategies most likely to reduce fatalities and injuries, and can also fund and test the effectiveness of new strategies. It routinely and aggressively shares its knowledge with the states. For example, several states adopted seat belt laws in the early 1980s, which allowed NHTSA to track how those laws affected fatalities. Its analysis revealed that state laws that allowed police to pull people over to check seat belt use resulted in higher seat belt usage and lower fatality rates than those that allowed police to check for seat belt use only when they stopped drivers for other reasons.[36]

In addition to looking across states to try to detect practices associated with safer driving conditions, NHTSA sometimes looks across time within a state after a state changes its policies and programs to assess the impact of the change. This type of analysis can be useful because it controls for the variations in background variables across states that might affect outcomes, such as the length of the motorcycle riding season, without necessitating data collection on every possibly relevant aspect of that variation. Because states have changed their laws pertaining to mandatory use of helmets by motorcycle riders so many times over the last 50 years, for example, NHTSA and other researchers (often with NHTSA grants) have a robust understanding of the effect of helmet laws on fatality rates and injury severity.[37]

Measurement, Mandates, and Money: The Case of Motorcycle Helmet Laws

NHTSA struggles with the question: How can it best use what it learns to reduce fatalities and injuries? Should it limit itself to the "bully pulpit of expertise," giving states information NHTSA collects about problems and viable solutions, hoping each state will have the good sense to use it? Or should the federal government mandate that states adopt strategies demonstrated to be more effective? This is a recurring question in the federalism debate. In recent years, NHTSA—often with direction from Congress—has experimented with a variety

of approaches that provide valuable lessons about how the federal government can use performance goals, often in the form of mandates to meet minimum standards or to adopt programs demonstrated to be effective, combined with mandates to measure. NHTSA's experience with motorcycle helmet laws demonstrates how federal measurement, mandates, and money—in combination and separately—can motivate improved state performance. It also illustrates how, as argued in Federalist Paper No. 51, "the different governments will control each other, at the same time that each will be controlled by itself."

The Highway Safety Act of 1966 (P.L. 89-564) required the federal government to set uniform standards for state highway safety programs. With evidence that riding motorcycles without helmets cost lives and exacerbated injuries, the federal government required states to adopt laws stipulating that all motorcycle riders wear helmets. States that failed to adopt such a law would lose 10 percent of their federal-aid highway construction funds.

The threat of a significant reduction in federal highway construction funds worked. Most states moved quickly to adopt universal helmet laws. In 1966, no state had a motorcycle helmet use law. By 1975, universal helmet laws had been adopted in all states but California and Utah.

In 1975, the Secretary of Transportation moved to exercise the penalty power in the 1966 law against the non-complying states. The states protested and Congress responded. It amended the law to prohibit federal imposition of penalties as well as the establishment of a federal standard pertaining to helmet use. Without the threat of a federal penalty, 28 states dropped their universal helmet laws.[38]

In the early 1990s, concerned about the continuing high number and cost of deaths associated with helmet-less motorcycle riders, the U.S. Congress, with leadership from former Senators Daniel P. Moynihan and John Chafee, again positioned the federal government to try to motivate changes in state law and rider behavior. This time, however, Congress took a less punitive motivational tack. The Intermodal Surface Transportation Efficiency Act of 1991 created an incentive grant program to reward states that had both a universal motorcycle helmet and passenger-vehicle safety belt use law. In addition, Congress added a potential penalty. Any state that failed to have a universal motorcycle helmet law and a safety belt use law by October 1, 1993, would have a portion of its highway construction funds transferred to its safety programs.

The incentives in the 1991 law proved less effective than the incentive of the 1966 law. Twenty-three states that already had both laws in place in 1991 received grant funds, but only Maryland and California adopted a universal helmet law between 1992 and 1995. In addition, the federal government transferred funds from state construction funds to the safety account in over half the states to penalize them for non-compliance.

States moved more quickly to change the 1991 law rather than their behavior. In 1995, Senator Ben Nighthorse Campbell, a motorcycle rider who liked neither helmets nor federal mandates on states, led the congressional effort to repeal the penalty. Following that repeal, four more states subsequently amended or repealed their universal helmet laws.[39]

The federal experience with universal motorcycle helmet laws clearly demonstrates the power of performance goals (mandates), measures, and money to improve societal outcomes. But it also demonstrates the vitality of the federalist system of checks and balances for gauging and addressing the outcomes that concern the American people.

Measurement of societal outcomes and tracking of program characteristics led to a policy that effectively reduced a large number of serious traffic-related injuries and fatalities. At the same time, resistance to helmet laws in many states reflects differences of opinion about how to balance two valued outcomes—greater safety and the joy of riding helmet-less. The federalist system afforded a voice to people with both points of view. The authority of both levels of government to set goals and measure progress toward them keeps each level of government responsive to its citizens.

Incentives and Marketing: The Case of Seat Belt Laws

Congress similarly created an incentive grant program in Fiscal Year 1999 to encourage states to

The Government Performance and Results Act, the Department of Transportation, and the States

As case studies two and three make clear, at least two DOT agencies and states have a long tradition of working together to measure outcomes and track different types of government intervention to identify the most serious problems needing attention, as well as successful solutions. Sometimes, the federal activity has been designed to assess and motivate progress toward a time-certain, quantity-specific target. More often, it has been used to compare across locations and times to identify effective government practices. Thus, when the Government Performance and Results Act (GPRA) passed, it did not require that either the FHWA or NHTSA make radical changes in their measurement practices. It did, however, change federal agency practices in subtle but significant ways. It also rekindled state fears about the possible exercise of unreasonable federal power.

GPRA Leads to Subtle but Significant Shift in Use of Performance Measures

When GPRA was passed, DOT's senior management aggressively pursued its implementation inside the agency, not just as a way to report to Congress, but as a core management mechanism. In an effort to support DOT's strong emphasis on outcome-focused management under GPRA, one NHTSA regional administrator changed the way he incorporated into his actions the outcome and effectiveness information the agency already gathered. He describes his transition:

> Prior to GPRA, states had to ask us for permission to do things, and we looked more at process than outcomes. If a state was going to buy a piece of equipment worth over $500, for example, it needed our approval. We still do some of that, but far less, and we don't expect states to come to us for approval. Instead, we analyze their performance and continually monitor what they are doing with regard to their performance and the legality of the purchases.
>
> Our jobs were more about process: Were the rules followed? We reviewed a lot of files and moved a lot of paper, not always equating what was going on to the bottom line: Are the roads safer because of our efforts? We did lots of management reviews based on an established timetable. Now we do management reviews if the grantee's numbers are not moving in the right direction.
>
> Now my job is: How can I motivate the states so they will do better? Before, we had the fatality numbers, but we didn't keep our eyes on the ball. Now, I will write a congratulatory letter to the state with the lowest fatality rate or the greatest improvement.

This region has hired a consultant to work with individual states to help them set specific outcome-focused performance goals and understand how to use the goals in managing their own programs. In addition, the regional administrator tries to inspire state improvements through face-to-face meetings and interstate comparisons.

GPRA Rekindles Fears about Accountability and Federal Power

The DOT effort to implement GPRA rekindled state concerns about how the federal government would use goals and measures in working with the states. Given that the DOT depends so heavily on the states to improve outcomes, particularly those of concern to the FHWA and NHTSA, is it even reasonable for the federal government to set specific outcome goals and targets? If so, who should set those targets? How ambitious should the targets be? And is it reasonable for Congress to expect the federal agency to achieve outcome targets it depends upon the states to meet? DOT and state agencies are struggling to find reasonable answers to these questions.

In an early post-GPRA attempt to strengthen the outcome orientation of the federal/state relationship, DOT proposed to link mandatory measurements to the award of formula grants. The states strongly objected to this idea, arguing that a link between federal performance measurement and state funding would restrain the flexibilities states had won through the 1991 ISTEA law. AASHTO issued a formal warning on the subject:

> Performance measures should not be used by the federal government as a means of restricting the authority and flexibility of state transportation officials, of complicating or further regulating the program, or creating additional data collection burdens on the states.

GPRA also stirred state fears that the federal government would expect states to meet federally set national goals and measure each state's progress toward the federal goals by generating a report card for each state. The states even questioned why, under GPRA, the federal government should set any federal outcome goals and measures since the federal government depended on the states to achieve the outcomes. As one NHTSA manager explained:

> We do not control fatalities on the highways. We influence the way the states spend their money and work with NHTSA. Lots of people in state transportation agencies object to our goals being so outcome-focused. We have had that issue pushed back to us from AASHTO and individual states.

Instead, the states proposed that the federal government measure federal performance relative only to its own activities: specifically, how quickly the federal government delivers appropriated money to the states and processes state requests.

DOT rejected this suggestion, opting to stay focused on the five priority outcome goals it had adopted: safety, mobility, national security, economic growth, and the human and natural environment. To allay state concerns, however, DOT agreed not to measure progress toward performance goals at the state level and to report only aggregate national results in its GPRA report.

In 1996, as one of a small number of pilot agencies testing early implementation of GPRA, NHTSA initiated an effort with 13 states to let them set their own safety goals. An evaluation of the pilot indicated that states took greater ownership of the goals when they set them and that focusing on outcome goals resulted in less paperwork. Unfortunately, the evaluation did not assess whether state "ownership" translated to performance gains.

Based on the pilot experience and influenced by GPRA, NHTSA opted to mandate state adoption of a performance-focused approach in 1998, hoping the mandate would change the behavior of both the states and the federal government. NHTSA is finding that state support for managing to meet outcome goals is growing with experience. State interactions with NHTSA are shifting to mitigating safety problems instead of complying with mandated processes. Federal managers feel the focus on outcomes is encouraging the federal government to work with the states in a more informative and less prescriptive manner, as well.

> We no longer say, "I am from the federal government and here is how you need to move your utility poles." Instead, we share with the states the leading causes of fatalities, and help them see how they could change the way they install utility poles to reduce fatalities.

Increasing state support for outcome-focused goals is also evident in a 1997 decision by AASHTO to adopt, for the first time, its own national outcome goal for the states. It adopted a goal for fatality reductions, neatly aligned to the federal government's priority objective of improved safety.

One challenge that looms large in the minds of federal agency staff as they move to implement GPRA is the difficulty of managing congressional and other external expectations regarding goals. At the Fiscal Year 2003 House Appropriations hearings, Congressman Rogers took NHTSA Administrator Runge to task for not meeting the agency's seat-belt and alcohol goals, despite the fact that the agency had aggressively encouraged and supported state efforts to adopt the changes needed to meet the goals. Runge's predecessors had adopted stretch goals for seat belt and alcohol use, aware of the possibility that the goals might not be met but mindful of research showing that a few focused stretch goals encourage greater innovation and motivation than more moderate goals. At the same hearing where Rogers reproached Runge, the head of the police chiefs in New York called the ambitious goals destructive. Runge subsequently revised the goals to be less ambitious, although DOT staff note that the targets were still high enough to place them among the leading contributors to DOT's overall safety goal. Again, the administrator came under attack, this time for backing away from the seat belt and alcohol efforts. This experience has left many in the agency gun-shy and wary

(continued on page 32)

(continued from page 31)

of setting all but readily achievable targets. It makes clear that if Congress and other external agency watchdogs hope federal agencies will innovate to increase their effectiveness and productivity, they may want to commend agencies willing to adopt focused stretch goals, rather than criticize them. It also suggests that federal agencies will need to experiment with their state and local delivery partners to sort out how to adopt shared stretch goals.

Looking Forward

DOT and the states are actively exploring sensible strategies for GPRA implementation across all modes of transportation. They have made great progress over the years, but still need to sort out many issues, especially as both the federal agencies and the states move to manage for results. This includes how to incorporate state-set goals into the measurement system and state-specific measurements into the national repository of experiential learning so the information gathered will benefit all states. In addition, the statistical validity and timeliness of the data periodically gets criticized and needs improvement.

Some of the approaches under consideration for the future include measuring federal progress toward the agency's outcome goals precisely, but measuring and reporting only the directional trend for each state; regional offices working with states to gather information about the goals they deem appropriate and toward which they could contribute; and more regional offices helping states benchmark with appropriately comparable states but without a full-scale 50-state comparison. Further, to respond to states' clearly articulated concern about their need for DOT to be more responsive, DOT is considering adding metrics on the federal government's speed in delivering grants to states after appropriation and its speed responding to state requests. One particularly acute state concern pertains to timeliness in federal action with respect to environmental processing.

increase seat belt use. The grant amount is based both on the state's level of performance relative to other states and its level of improvement relative to its own performance.[40] This provision has had an interesting effect on the states' attitude toward measurement. Since the pot of reward funds to be shared among the states was finite, the states pushed to assure fair performance measurement through federal standardization and quality control of the measurement methodology.

NHTSA not only analyzes its data to understand the primary sources of problems and identify effective solutions. It takes great care to market the knowledge it has acquired about the relative severity of problems and the availability of effective solutions to key customers, most notably the state and local officials who can help it accomplish its objectives. This is especially important when it lacks the power it had with the early motorcycle helmet law of a strong regulatory hammer. The absence of that power necessitates other federal strategies. NHTSA disseminates information in a manner that is easy to find and easy to use. It develops and aggressively promotes campaign materials that other levels of government and NGOs can use. It packages campaigns but also steps back to understand the needs of individual states, and work with them to find effective implementation strategies that take local attitudes and authorities into account.

NHTSA's website is exceptionally customer focused, describing the problems it addresses in ways likely to help others achieve increased traffic safety.[41] The site contains reams of easy-to-understand, well-packaged material explaining key contributors to traffic safety problems, costs associated with traffic accidents, and evidence about the effectiveness of efforts to address the problem.

NHTSA also runs several outreach campaigns to help states increase seat belt use. This includes the "Buckle Up America" campaign, providing states and others with well-developed marketing material and plans to help them increase seat belt use. As part of that effort and based on evidence that education alone is insufficient to raise belt use levels beyond the 65 to 70 percent level, NHTSA and states are testing and evaluating a carefully designed campaign that combines public relations,

paid marketing, and enforcement. The program has been dubbed "Click It or Ticket." Using a control group of states that did not implement the program, NHTSA credits a multi-state Memorial Day blitz using the "Click It or Ticket" program with a 2 percent increase in belt usage.[42]

Public tolerance for seat belt laws and random checks for seat belt use can vary significantly by state, just as people in some states tend to drink more than those in others, independent of any explanatory government actions. Recognizing these differences, NHTSA regional and state administrators work closely with state leaders to craft location-acceptable strategies. The "Click It or Ticket" program, for example, was modeled after a stop-and-check effort tried in North Carolina based upon a Canadian effort. South Carolina wanted to replicate it, but lacked North Carolina's legal authority to stop and check. The regional administrator worked closely with South Carolina and other states to adapt the strategy. As a result of the campaign, belt usage in the entire region increased an average of 9 percent.

Coaching and Competitive Leagues

Another interesting aspect of NHTSA's implementation of outcome-focused management and measurement is the unofficial coaching role a few regional administrators have chosen to play. One NHTSA regional administrator is experimenting with ways to create an atmosphere of friendly competition for his states that is motivating, but not debilitating. He distributes to states in his region an informal inter-state analysis. The regional administrator explains:

> I try to create competition with the states. I have prepared a spreadsheet comparing the states, and I send it out to the states in my region. There is no federal rule requiring that, but I do it and send it out. Most of the time, the states buy into the information and call each other to ask questions.

Limiting the comparison to a subset of states makes the volume of the data more manageable. By narrowing interstate comparison to states in a single region and bringing states' attention to the comparisons, this NHTSA manager focuses attention on states more likely to share cultural and economic characteristics. Neighboring states may also be easier to study because of their proximity and because they sometimes share a media market. Some regional comparisons may tap into age-old local rivalries. Similar to youth sports leagues, where skill level or regional segregation makes winning more attainable, limiting state comparisons to a subset of states—thereby increasing the number of states in the country who can be at the head of the pack—may strengthen state aspirations for advancement.[43]

Using data compiled nationally, the regional administrator also compares states in his region to the national front-runners to search for relevant lessons. It is interesting to note how FHWA headquarters supports this regional effort. Analysis included in the annual *Highway Statistics* report on "Selected Measures for Identifying Peer States" makes this sort of benchmarking more feasible.[44]

Another federal Transportation manager describes his dream for how the federal government should partner with states to improve outcomes:

> We can create communities of practice. In the field, we [in the federal government] can get our head up above the daily work to look around and see what is working and to look for new technology that looks promising. States are really overwhelmed just by getting the job done, dealing with the contracts, handling the environmental issues. They don't have enough time to look up and around and find how to do things more efficiently. We can do that at the federal level to help them.

Case Study Four: Department of Education—Use of Comparisons to Leverage National Goals

Although the U.S. Department of Education is one of the youngest federal cabinet departments, the federal government has long collected state and local education measurements. Half a century before the federal roads office began collecting data on local road conditions, federal officials began collecting and publishing data about schools, attendance, staffing, finance, and transportation. Until the 1950s, however, the federal government gathered data only at the state level, essentially taking whatever data the states were willing to provide. Since variations in education practices primarily occurred at the local level, the absence of local data precluded analysis to look for the variations associated with different performance levels. Also, the lack of common measurement methods across states limited its interpretability.[45]

In 2001, the U.S. Congress, with strong encouragement from former governor and President George W. Bush and leadership from Senator Edward Kennedy, enacted a new education law, the No Child Left Behind Act of 2001 (NCLB).[46] This law wholeheartedly endorses the principle that the federal government should play an activist role encouraging the use of outcome-focused performance measurement by states. It does not, however, call for direct federal measurement of state performance.

NCLB strengthens mandates established in the 1994 federal education law[47] that states establish rigorous performance standards for all students, measure progress toward them, and hold schools accountable for performance progress. The 1994 law required states to comply with these requirements by January 2001, allowing the U.S. Department of Education to grant a waiver extending the deadline for states making adequate progress coming into compliance, while authorizing the department to withhold federal funds for administering Title I of the Elementary and Secondary Education Act from states that are not. By March 2002, 17 states had complied with the accountability and assessment requirements of the law. The remainder had not.[48]

NCLB continues the performance measurement requirements established by the 1994 law, including the sanctions for non-compliance. It also adds a requirement—that states annually test all students in grades 3 through 8 to measure progress toward the standards. The new law clearly rejects the idea of a uniform set of federal educational performance measures, preferring instead that states and their citizens establish their own measurement systems reflecting state, not federal, preferences regarding educational curricular goals.

The law does, however, strengthen an existing national measurement system, the National Assessment of Educational Progress (NAEP). In addition to conducting their own annual performance testing for all students in grades 3 through 8, states are required by NCLB to participate in NAEP math and language-arts testing in grades 4 and 8. Enhancing NAEP testing enables cross-state benchmarking. NAEP allows researchers to normalize across state measurement systems, so states, their citizens, and the federal government can assess the rigor of each state's standards. The new law prohibits the use of NAEP data, however, as the

basis for sanctions on individual teachers, students, schools, or states.

In the years to come, the No Child Left Behind Act will undoubtedly offer many valuable lessons regarding effective and ineffective ways the federal government can use performance goals and measures to assist states in improving program results. Past experience offers instructive lessons as well. Several vignettes are presented here. The next section provides a brief history of federal handling of state and local education measurement to provide a context for understanding the other two examples presented as promising exceptions to the conventional way the federal government has approached state and local educational performance measurement. The two examples—the release of the "Wall Chart" in 1984 and, more recently, reports released by the Education Trust, a nonprofit organization, that use data assembled by the U.S. Department of Education—illustrate the power of comparative measurement across jurisdictions whether within a state, among states, or internationally. The wall chart also underscores the power of state political leaders collectively endorsing comparative measurement, while the Education Trust example illustrates the tremendous value that can be reaped when the federal government simply gathers and organizes state performance data and makes it readily accessible for others to use.

The Conventional Approach to Education Measurement

Various efforts have been made since the early 1950s to enhance the value of the education measurement system. In 1954, federal and state education statisticians collaborated in the development of a common set of definitions. Congress created the National Center for Education Statistics (NCES) in 1965, which attempted the first national analysis of local school districts using statistical sampling. In the late 1970s, the National Center for Education Statistics within the federal education agency initiated what is known as the Common Core of Data (CCD). The CCD represented the federal government's first concerted effort to assemble a robust and regularly collected set of education data from local school districts and states to inform managers and policy setters.[49]

Unfortunately, as social policy experts Weiss and Gruber report in their 1987 essay tellingly titled "Managed Irrelevance of Federal Education Statistics," federally collected education statistics have historically proved of limited value to government policy makers and managers. The bulk of the information the federal government has collected from states and localities has been gathered to serve the needs of federal program managers seeking to assure that grantees spent their federal funds as directed; the program offices that controlled the collected data made little effort over the years to organize it in a manner that could help others learn from the experience gained through those expenditures.

Until the end of the 20th century, the value of the CCD effort was severely hampered by the politics of intergovernmental and intra-governmental performance measurement. Few rallied to support the effort to build a strong national database of education statistics that would allow educators, education managers, parents, and elected officials to find successes worthy of replication and less successful strategies in need of curtailment. In contrast, an abundance of well-armed opponents routinely rallied in opposition.

Among the leading opponents were state and local educators and education administrators who saw little to be gained and much to be lost by having the federal government assemble state and locally generated education measurements. In their essay, Weiss and Gruber succinctly summarize the barriers to creating a federal repository of useful state and local measures. This problem was undoubtedly exacerbated by the fact that, in contrast to the federal road agency whose existence preceded all but two of the state road agencies, state governments had long managed and had their own agendas for their education measurement activities.

> The trouble is that the federal objective of [data] standardization competes with the dominant state technical objective for federal data collection, which is to limit the work required. The people in the states who provide data for CCD spend most of their time collecting data from local school districts to comply with sundry state mandates. … the CCD and [National Center for

> Education Statistics] do not loom large in their lives. State and local procedures do.... They are up to their ears in state-mandated work and local intransigencies about reporting. They do not have the slack to accommodate changes in definitions and collection schedules to suit the convenience of some bureaucrat in Washington…. To over-generalize only slightly, their agenda can be summed up as "no changes."…[50]

But practical issues are not the only impediment. Political positions are an equally, arguably more, important factor.

> The gap between state and federal perspectives is also apparent when it comes to political agendas. The major political objectives for statistics at the federal level are to inform Congress, the White House, and top executive officials about national developments in education, to allow comparisons among the states that may point up the problems or progress, and to develop a respectable national archive of data that satisfies research, planning, and forecasting needs in and out of government.…
>
> But the states' agenda is just the reverse. State officials such as the Council of Chief State School Officers seek to limit outside intrusion into what they regard as state business…. As state people told us, the more federal officials and Congress know about what is going on in the schools of Ohio or Arizona, the more likely they are to try to tinker. And, if the federal government makes available public data about how Ohio and Arizona compare, that provides fodder for disaffected state legislators or irate taxpayers to make life difficult for state education people.…
>
> In line with this pattern, the states have not opposed all federal education statistics, only those that increase their vulnerability to external attack. Indeed, CCD in its present form enjoys substantial moral, logistical, and political support from state officials.… CCD's principal attraction[s] to state data collectors are the ability to make realistic comparisons to other states on questions of size and resources (but not performance), and the implicit endorsement and explicit assistance of the federal government and other states for improving the technical quality of the data (and thereby resisting local political pressure to manipulate the data).[51]

The lack of a federal repository that contains state and local measurements of educational outcomes along with data about activities and resources has seriously constrained the possibility of learning from experience. Many blame the perennial disagreement about the appropriate social objectives for education programs for the lack of adequate outcome information. Should schools simply teach reading and mathematics, prepare students to be good citizens in American society, or ready students for the workforce? Should educators strive for functioning or flourishing graduates? Do our schools need to address social harmony, social equity, or the nation's economic vitality? These questions are greatly complicated by heated debate about the need for, and appropriateness of, any federal role in the education effort. The primary emphasis of federal education policy, until the passage of NCLB, was educational advancement for the disadvantaged, not educational advancement for all. Still, the absence of outcome information relative to any of these goals does not advance one at the expense of another; it simply means that the debates about objectives that take place occur without information about the likelihood of programs advancing any one of them.

The private sector has filled some parts of the data gap. The College Entrance Examination Board, for example, tests college-bound students and the National Education Association collects information about teacher working conditions. For the most part, though, no sustained effort was made to fill the gap in education measurement, especially outcome measurement, for several decades. The prospects for more useful educational measurement did not begin to improve markedly until the end of the 20th century when two provocative federal agency leaders used comparative performance measurement to galvanize attention to educational

problems, and the nation's governors, two of whom subsequently became president of the United States, responded.

The "Education Wall Chart"[52]

In 1984, the U.S. Department of Education (ED) published a large one-page chart comparing educational performance across the states. The comparative chart, which soon came to be known as the "wall chart," culled data from both public and private sources. It was created at the instruction of then-Secretary of Education Terrell H. Bell. The wall chart compared the performance of public schools in every state using SAT and ACT scores.[53] It also compared state-reported graduation and drop-out rates. In addition, it reported comparative inputs such as per-pupil expenditures and average teacher salaries, along with information about the background characteristics of students, such as the percentage in poverty and the educational attainment of parents.

The idea of comparative school performance measurement was not new. Horace Mann advocated it in the 1840s for the schools of Massachusetts.[54] Yet the wall chart represented, in the words of one of its creators, "the first ever state-by-state national report card on educational performance."

Comparison Motivates

Prior to creating the wall chart, Secretary Bell convened the National Commission on Excellence in Education. In 1983, the Commission released its report, *A Nation at Risk: The Imperative for Educational Reform*. Just as the wall chart would compare educational performance in the states to motivate improvements, the *Nation at Risk* report turned to international comparisons to motivate U.S. action on education. The report sounded a warning:

> International comparisons of student achievement, completed a decade ago, reveal that on 19 academic tests American students were never first or second and, in comparison with other industrialized nations, were last seven times.[55]

These and other findings in the *Nation at Risk* report galvanized leaders across the nation, especially governors and the business community, to action.

Among the recommendations in the *Nation at Risk* was a proposal for a nationwide system of educational performance measurement:

> Standardized tests of achievement (not to be confused with aptitude tests) should be administered at major transition points from one level of schooling to another and particularly from high school to college or work. The purposes of these tests would be to: (a) certify the student's credentials; (b) identify the need for remedial intervention; and (c) identify the opportunity for advanced or accelerated work. The tests should be administered as part of a nationwide (but not Federal) system of State and local standardized tests. This system should include other diagnostic procedures that assist teachers and students to evaluate student progress.

Gubernatorial Leadership Catalyzes Inter-State Comparisons

For years, states and professional associations of teachers and school administrators had resisted comparative measurement. The *Nation at Risk* report, together with gubernatorial and agency leadership, proved the necessary catalyst to surmount this resistance. Concerned about the findings of the *Nation at Risk* report, governors turned to Bell at the 1984 meeting of the National Governors Association for assistance in assessing the performance of schools in their own states. Bell readily agreed, producing the wall chart.

Not surprisingly, the release of the wall chart garnered great public attention. Over 30 television networks and local newspapers across the country covered the initial press conference. To put educational improvement on each state's agenda, Secretary Bell and his successor, William H. Bennett, followed up on the press conference by traveling across the country promoting the wall chart. Educational performance became an election year issue in several 1986 gubernatorial elections, and the U.S. Department of Education discovered what most states feared: Federal release of compar-

ative performance measurement can become a popular local news story, often a political one.

Problems with Non-Standardized Performance Data Across States

Several aspects of the wall chart triggered criticism. Differences in the percentage of students in each state taking the SATs or the ACTs skewed the results. Also, SAT and ACT scores captured only the performance of college-bound students. The wall chart results were less relevant with regard to students who did not plan to attend college after high school. Subsequent versions of the wall chart addressed some of the methodological problems. Problems remained, however, because nationally standardized state-level data were unavailable. This methodological weakness begged for a federal response.

In 1987, a study commission headed by a governor, Lamar Alexander of Tennessee, recommended just such a federal response, calling for the expansion of the National Assessment of Educational Progress to allow for state-by-state comparisons. In the same year, then-Governor Bill Clinton of Arkansas served simultaneously as chair of the National Governors Association (NGA) and the Education Commission of the States. Education was at the top of many governors' agendas, and the governors served as a driving and bipartisan force for national action on what had previously been regarded as a state and local policy issue.

Good Measurement and Analysis Promotes More Good Measurement and Analysis

The wall chart convinced many states long reluctant to measure educational outcomes to improve their own measurement efforts. Almost immediately, a few communities adopted the federal government method's for calculating drop-out rates developed for the wall chart because it provided a more accurate, albeit less flattering, picture than the locally generated ones. In 1986, eight Southern states paid NAEP for supplemental state testing to help them assess their own performance. This action was remarkable because Southern states had previously been the leading opponents fighting state-specific NAEP testing. Indeed, gun-shy about grading the states, Congress had prohibited the Department of Education from conducting state-based NAEP testing. Responding to evolving state views about the value of comparative educational testing, Congress finally authorized voluntary state testing under NAEP in 1988.

The wall chart jump-started educational performance measurement around the country. By 1988, 12 states had initiated comparative performance measures for their own local school districts. Today, only one state, Iowa, lacks any sort of uniform statewide testing program or plans to create one.[56]

Comparative Input Data Important for Identifying Performance-Improving Solutions

While the wall chart compelled states to face their problems, it did not provide much help in identifying ways to fix them. Using available data on per-pupil expenditures, average teacher salaries, pupil-teacher ratios, and an indicator of each state's fiscal effort for education, the wall chart provided a rough assessment of the effect of state fiscal and staffing decisions on outcomes.

The ability to find other likely causes of performance variations was limited, however, because data on other attributes of educational programs likely to influence outcomes were not available. In fact, in creating the wall chart, the Department of Education discovered it had stopped collecting data it had once collected on state program characteristics. To find teacher salaries for the initial chart, for example, it had to use data from the teachers' union. Creating the wall chart convinced the agency to reconsider its data collection decisions. In addition, several states, in setting up their own measurement systems, opted to enhance what they learned from the wall chart; they began to collect other input variables that might affect educational outcomes, such as amount of homework, courses taken, and attendance levels.

Lack of Legislative Mandate Threatens Sustainability of Measurement Systems

Unfortunately, William Bennett's successor opted to discontinue production of the controversial wall chart. As so often happens when comparative measures have been generated without legislative mandate or a strong and well-organized audience, opponents of measurement—those being measured—succeeded in killing the comparative measurement effort.[57] Moreover, few champions of the

wall chart resided among the program managers in the Department of Education. They saw the chart as a product of the "front office," not a critical tool to help them help states and localities find better methods for achieving educational objectives or even a tool for motivating improvements through incentives.

Despite its demise, the wall chart demonstrated how, by the simple act of gathering and disseminating state performance data, the federal government was able to motivate states to measure and improve outcomes.

The Education Trust: Third-Party Analyses

The wall chart demonstrated that federal analysis and dissemination of information can motivate measurement, contributing to performance improvement. More recently, a report released by the non-profit group Education Trust demonstrated that simply by assembling state performance data into a single database to make it easily accessible for study, the federal government can invite and support value-adding analysis by others.

Studying Success for Replication

In December 2001, a nonprofit organization, the Education Trust, released a report, *Dispelling the Myth Revisited*.[58] The report identified nearly 5,000 high-performing schools with high percentages of low-income or minority students. The Education Trust dubbed these schools "high-flyers." The Trust's goal in releasing the report was to dispel an assumption voiced by many public education leaders that schools teaching poor and minority children could not achieve strong academic performance because they were constrained by the societal difficulties plaguing the student population.

The 2001 report was the Education Trust's second effort to dispel low expectations policy makers and educators held for children in low-income and high-minority neighborhoods. The Trust conducted its first study in 1998 with cooperation from the Council of Chief State School Officers and funding from the U.S. Department of Education. For its 1998 study, the Education Trust asked the chief state school officers in every state to identify their top-performing high-poverty schools. They received responses from 21 states and contacted a thousand schools. The Trust "mined the data" to identify common characteristics of successful schools. Based on several hundred responses from those schools, the Education Trust was able to develop a preliminary list of characteristics contributing to success. They found that one of the most common characteristics of successful schools is the extensive use of state standards to design curriculum and instruction, assess student work, and evaluate teachers.[59]

Because of the limits of their methodology (each state defined top-performing schools in its own ways), Education Trust Director Katie Haycock noted in the preface to the 1998 report, "Future reports will, we hope, include more complete information on some of the schools, as well as information from top performing schools in the states that did not participate in the initial study." Three years later, the Education Trust fulfilled its own hopes, releasing a preliminary report identifying 4,755 high-flying schools across the country.

For the 2001 report, the Education Trust refined its definition of high-achievers. Rather than looking at the top-performing schools relative to other high-poverty schools, it looked at schools with high achievement (top third) relative to all other schools in the state based on statewide testing. Within that subset, it identified schools with greater than 50 percent Latino and African-American populations, schools with greater than 50 percent low-income students, and schools with both.

Value of Federal Government as Information Assembler

The ability of the Education Trust to conduct its more complete and accurate 2001 study was greatly facilitated by a decision by the U.S. Department of Education to fund the creation of a nationwide database combining school-level assessment scores from all states as part of its effort to analyze the correlation between NAEP and state tests. State policy decisions combined with technology advances facilitated the compilation of the data. By 2000, at least 30 states had opted to post

school-based data on the Internet. The department turned to a contractor, the American Institutes for Research (AIR), to harvest school-performance measurement information states posted on the Internet, supplement it with educational performance information gathered from the other states, and create a single database. The AIR database also includes other basic school characteristic information the department collects, including fiscal, program activity, and demographic data.[60] By providing a single portal to comparative state outcome and input information, the U.S. Department of Education has greatly reduced the costs and increased the prospects for problem and solution-finding analysis.

Encouraging Audience-Focused Analysis

The department's decision to gather, organize, and facilitate access to information from the states invites external organizations, such as the Education Trust, to select the subset of data that answers their own questions, which may be different from the questions of the department or other organizations. Creation of an accessible national repository of performance information—including outcome, output, and input data—engages the resources, perspectives, and insights of external parties. The Education Trust conducted analysis to "dispel the myth" that poor and minority children are necessarily poor performers. It has subsequently released reports showing the disproportionate numbers of classes in high-poverty and high-minority secondary schools taught by teachers lacking a major or minor in the subject they teach and the funding differences for schools with different income and racial characteristics.

GreatSchools.net uses the same data to provide an online guide for parents of all incomes to inform their housing purchases or simply to learn more about the comparative performance of their children's school. By creating an information repository of performance and input data that others can easily access and analyze, the U.S. Department of Education facilitates audience-focused, performance-improving analysis of state performance data. This can introduce market-like mechanisms using performance measures to inform citizen choices when they directly consume government-provided or government-regulated goods. In other cases, it informs citizen decisions about their elected officials. This, in turn, motivates and helps state leaders to improve students' educational performance.

Good Federal Infrastructure Promotes More Good Infrastructure

One further aspect of the Education Trust 2001 analysis merits mention. In releasing *Dispelling the Myth Revisited,* the Education Trust simultaneously announced the creation of its own enhanced version of the U.S. Department of Education's AIR database.[61] In doing so, the Trust clearly hopes to engage others in using the data:

> We hope others will join us in the endeavor by using a new tool on the Education Trust's Web site…. This site allows users to search for high-performing schools according to users' own achievements and demographic criteria.

Both the department and the Education Trust realize that building an easy-to-use gateway to comparable and relevant information enlists the brains and efforts of many external experts, leading to improved societal outcomes.

Accountability-Assuring vs. Information-Improving Information

The U.S. Department of Education conducts its own analysis of school performance data as well, but while the Education Trust concentrates on high-performing schools, the department looks at low-performing schools. Both are informative, but serve different purposes. The federal government uses the performance data to find problems and assure state accountability for fixing the problems. The Education Trust uses it to find and share successes worthy of replication.

Federal law has long focused on weaknesses in the educational system. The Elementary and Secondary Education Act of 1965 directed federal attention to students with background characteristics seen as being associated with low performance, including economically disadvantaged students, non-English speakers, Native Americans, and migrants. The federal government provided financial assistance to states to encourage them to help these needier students. The 1994 law, adopted after states began to rely more heavily on performance measurement,

changed the orientation of the federal law slightly. Title I of the 1994 reauthorization of the Elementary and Secondary Education Act required states and school districts to identify low-performing schools, using rigorous state-set standards of education applicable to all students. It also required states to assist those schools, with the federal government withholding funds from states that failed to do so.

States moved slowly to implement this aspect of the 1994 law. As of January 2001, seven years after the law established this requirement, only 27 states had a policy in place to identify the low-performing schools, let alone intervene in them. Frustrated by slow state action but reluctant to withhold funds precipitously lest it trigger state efforts to revoke the sanction authority, President Clinton issued an Executive Order directing the U.S. Department of Education to report on each state's progress identifying low-performing schools needing assistance.[62] The first such report was issued in 2001.

An interesting aspect of the Executive Order is that it focused on states' ability to find problems, but ignored their ability to find solutions to those problems. The report noted the need to find successful interventions, lamenting the scarcity of such analysis:

> Research on effective schools is plentiful and largely in agreement in its conclusions. Research on the process of turning a low-performing school into an effective school is much less plentiful and more difficult to interpret….[63]

The department report cites the 1999 Education Trust success-finding analysis and the reports of two other nonprofit organizations that have begun the search for replicable patterns of success.

That the Education Trust, an external advocacy group seeking to improve education, chose to conduct and release a success-finding report should not be taken for granted. Many external advocacy organizations opt instead to focus only on negative state performance, lambasting poor performers and creating fear and resistance about comparative federal performance measurement. In order to obtain access to the data compiled by the American Institutes for Research, the Education Trust had to assure states that it would not produce a report making the states look bad. Still, by offering constructive rather than critical analysis and helping schools in their own struggle to find solutions, the Education Trust models an alternative use of measures that can enhance their long-term viability because those being measured will not see the measurements only as ammunition for attack. If external advocacy organizations use performance information to praise those doing well, and to help those doing less well find viable alternative approaches to the problems they must address, they may enhance the survival prospects and usefulness of outcome-focused performance measurement.

Surmounting Political Constraints Curbing Honest Analysis

As noted earlier, in stark contrast to the work of the National Highway Traffic Safety Administration, the Department of Education has not historically emphasized the search for and promotion of successful government interventions. By its legislation and leadership, it has played a problem-finding, accountability-assuring role. Its forays into success finding and best-practice promoting have been more incidental and case focused than systematic.

One contemporary education expert external to the department suggests that politics constrain the department's ability to do an honest assessment of different educational programs:

> The government seems reluctant to report performance in areas federal dollars are spent. The U.S. Department of Education has tried on and off to report on the quality of programs and the quality of schools, but in my experience, the federal government has not had the capacity or it was fearful of provoking controversy by making program developers in some congressperson's district unhappy. So for the most part, the department has not been able to do this.

The expert goes on to recommend a solution:

> The right thing for the feds to do is not to get into the rating business, but to allocate funds so independent organizations can

conduct the analysis. This would insulate the government from findings that have consequences for people in terms of policy and revenue streams.

The media's response to the release of the Education Trust's report suggests this may be a good strategy for some federal agencies. The media gave the Education Trust report significant attention, reflecting favorably on both the department and the Trust. A recently issued request for proposals from the department's Office of Educational Research and Improvement for a "What Works Clearinghouse" to create a registry of demonstrably effective practices is a promising sign that the department will find politically viable ways to build its success-finding capacity and create its own "bully pulpit of expertise."

Armed with a new law calling for a stronger federal role in using state performance measures to improve the education of America's children, with new authorities for the department and automatic triggers of consequences linked to low performance (e.g, the law entitles children in perennially low-performing schools to change schools and requires states to contribute to transportation costs for those children), the current leaders of the department have indicated their intentions to integrate performance information collected from states and localities into the heart of the department's operations. They have already taken a noteworthy first step, posting on the Internet the number of perennially low-performing schools in each state where students can opt into other schools and receive state transportation support.[64] It is hoped that, in the future, the department may resurrect a more sophisticated "wall chart"-like analysis, whether it does it itself or funds others to do it.

Summary

In sum, the experience of the U.S. Department of Education illustrates the power of effective federal handling of state and local performance measurement. The wall chart and the *Nation at Risk* report demonstrate how comparative measurement can motivate improved performance. The American Institutes for Research database shows how, just by gathering and organizing information and making it easy to use, the federal government can stimulate external analyses to find effective government programs worthy of replication. At the same time, the difficult history of federal efforts to measure educational performance is a powerful reminder that the federal government must make it a priority to translate the information it gathers into products that those being measured and citizens will value highly. It also demonstrate the powerful contribution state and local political leaders can make collectively when they move beyond their fear of finding problems in their own systems to endorse the adoption of comparative measurement.

Findings and Recommendations

The federal government has long used measurements, mandates, and money in its work with the states to encourage greater social results. The success of its efforts depends on how skillfully it balances these three tools.

Findings

Finding 1: Analyzing Measures Adds Value

As a general rule, federal agencies that have emphasized and publicized the informational value of state performance measures—for finding successful practices, identifying significant problems or patterns of problems, and triggering focused follow-up inquiries to find key contributors to outcome variations—have been able to build more sustainable and valuable measurement systems than measurement systems used primarily to assure the fulfillment of commitments made or mandated. Common measurements across states and across time are useful for identifying problems to be addressed and successes to be replicated. They are also powerful motivators when accepted by those being measured as legitimate or when the public wants to use the measurements to inform their own choices. By its actions and the way it uses performance information, the federal government can encourage the creation and maintenance of a learning environment focused on harvesting the insights and motivational potential of accurate and comparable state performance measurement systems.

When common measurement systems do not already exist, congressional mandates to measure state outcomes and activities and report them to the federal government are often necessary. Federal funds attached to measurement mandates make them more politically palatable. Even in the absence of such a mandate, however, federal agencies can often work with states to develop common measurements. And if federal agencies fail to step forward to create a common measurement framework, state agencies can work together to create one. Federal failure to use and share state data, especially with those who supply it but with others as well, squanders its potential for improving program effectiveness and efficiency. It can also render mandated state measurement a meaningless exercise and result in states paying little attention to the quality of the data they submit, because they expect so little in return.

Finding 2: Federally Mandated Goals Work, but Can Be Problematic

While often politically controversial at the state level, federally mandated goals for states can be powerful motivators when linked to the promise of significant rewards or the threat of significant penalties. The promise of a reward or the threat of serious penalties linked to goal attainment can add to the motivational value of measures as long as those being measured do not feel so strongly threatened that they try to have the goals repealed or destroy the measurement system, either by dismantling it or by undermining it with inaccurate and untimely measurement submission.

Recommendations

1. Collect, Organize, and Make Information Readily Available
2. Create Robust Measurement Systems
3. Standardize and Normalize
4. Require Measurement
5. Involve and Benefit Those Being Measured
6. Encourage Analysis
7. "Market" the Results
8. Motivate with Comparison and Rewards, but Carefully
9. Share Best Practices

To lessen the chances that states will organize to dismantle the measurement system, federal agencies should make it a priority to build measurement systems that serve the needs of those they measure, most notably their state and local government service delivery partners, along with others whose actions contribute to improved outcomes and who face choices among options that might be influenced by the performance information.

Recommendations to Improve Performance Goals and Measures

The following findings and recommendations are intended to catalyze conversation in Congress, federal agencies, the states and their trade associations, and public interest groups about the question, "How can the federal government best use performance goals and measures to work with states to improve societal outcomes?" Beyond debate, it is hoped these recommendations will inspire readers to move forward to experiment with and implement better ways to tap the power of performance measurement. It is time to strengthen federal agency capacity and understanding to work with states and others to use performance goals and measures more effectively to achieve better program results. With the new Department of Homeland Security in mind, the federal government should recognize the importance of building a measurement system that will serve not only the federal agency but, critically important, the state and local governments that carry out much of the nation's daily security-enhancing activities.

Recommendation 1: Collect, Organize, and Make Information Readily Available

Federal agencies can play a valuable role few others can play simply by gathering performance information from each state, organizing it to facilitate interpretation, and making it broadly available. The ready availability of easily interpretable information about the full spectrum of performance—including information about outcomes, outputs, activities, inputs, and relevant background characteristics—enables the search for what works and what does not. Making it easier for states to learn what other states are doing adds value to the information, especially when the information is updated annually. It is especially important to return the information back to the data suppliers in a format that adds value compared to when it was submitted. Failure to share performance information that is gathered converts the measurement effort into a meaningless exercise and a waste of resources. It also frustrates those being measured who fear the comparisons.

For example, by gathering, organizing, and annually publishing data about the characteristics and conditions of U.S. roads in the annual *Highway Statistics* report, the Federal Highway Administration has long helped states learn from each other's successes and problems. Despite the lack of a congressional mandate for states to supply the data or a formal sanction process for failure to supply it, states have historically supplied data to FHWA because the agency makes it a priority to return the collected data back to the states in a format that gives it greater value than when the states supplied the data.

Actions to Implement Recommendation 1

- **Federal agencies** should make it a top priority to collect and organize state performance information in a readily accessible, easy-to-use data repository. Federal agencies should do this in all cases for federally collected data related to the agency's outcome targets, but they should also include information about relevant data from other public and private sources that is of decent quality. This will facilitate its use by others who share the agency's objectives and are similarly searching for effective ways to make progress toward those objectives. Agencies should organize this information to facilitate interpretation and analysis, not only by making the data from each state available, but also by organizing it by category across all states.
- Each **federal agency** should make clear in its annual GPRA performance report what relevant, state-specific performance information is available and where to find it. Congress should require federal agencies that work with states to publish annual (or biannual) compendia of state-specific performance information. It should provide funding and training to state staff for that purpose.
- **State associations** should support and work with federal agencies in the effort to collect and analyze useful state performance and other information.
- **OMB and GAO,** which work with multiple federal agencies, should facilitate learning across the federal government on the best ways to handle state information to advance outcomes.

Recommendation 2: Create Robust Measurement Systems

When federal agencies gather and organize data from states about all dimensions of their program—including information about outcomes, outputs, activities, inputs, and relevant background characteristics—it supports the search for what works and what does not. It also helps the federal agency present more informed arguments about resource needs to federal and state decision makers. Federal agencies already gather much of this data for program management purposes, but do not always organize and disseminate it in a manner that facilitates access and interpretation.

For example, NHTSA gathers data about fatalities and other accidents, the consequences of those accidents, and the likely causes. NHTSA organizes much of its data in reports focused on specific outcomes of concern, such as traffic crashes, or on programs identified through research as being effective for advancing its objectives, such as seat belt use. These reports are supported by well-documented analysis of state experience, with supporting data, and ready access though the Internet to the underlying databases.

Likewise, the decision by the Department of Education to fund the American Institutes for

Actions to Implement Recommendation 2

- For maximum interpretation value, **federal agencies** should include in their data repositories the full spectrum of information from states—about outcomes, intermediate outcomes, outputs, and the inputs of state programs, as well as the characteristics of the people, places, and things being affected by or potentially affecting program outcomes. Federal agencies should do this in a manner that is easy to find and easy to use.
- Where federal agencies lack the full spectrum of state measurement data, they should take the necessary actions to supplement them, including seeking **congressional** support and mandates. **State leaders** and their **associations** should move past their fears and work with the agencies to craft workable proposals for congressional mandates to create useful compendia of state performance information. States should also work with federal agencies to build repositories of state data even in the absence of mandates.
- In some cases, **federal agencies** may want to fund third parties to create state databases. For example, EPA might want to support external researchers to develop an easy-to-use, easy-to-find database, comparable to that which the Department of Education funded the American Institutes for Research to create, containing annual data about ambient conditions, environmental discharges (including releases and emissions), compliance rates, program characteristics, and program inputs for all states, such as state expenditure and staffing data.

Research to gather outcome information from state performance tests and combine it with the output and input data in the Core of Common Data has greatly enriched the informational value of both sets of data for educators, policy makers, and the researchers who seek to support their decisions.

Recommendation 3: Standardize and Normalize

Data from states are much more useful if they are standardized and normalized—that is, information submitted from each state for a single data category has the same meaning and is characterized in units that enable appropriate comparisons. Standardization and normalization allows comparisons, which can motivate, inform choices among options, and identify strong performers with likely lessons for others.

States can play a leadership role initiating standardization when federal agencies have not focused on the issue. Federal agencies can support that effort by staffing, supplying information to, or serving as the secretariat for state-led efforts to set standards. They can further speed data standardization by inviting state-offered proposals as the basis for proposed federal regulations, which the agencies can then put through the normal public comment and review process. State data standardization efforts are likely to be greatly strengthened if they engage NGOs and other interested parties in the process, even before the formal federal comment period.

When neither the federal government nor the states carry out the needed normalization, NGOs can play a catalytic role, conducting analyses that can serve as a model for future federal or state efforts.

For example, federal and state education statisticians first collaborated on the development of common education metrics in the 1950s, but could not at that time nor any time until 2000 agree on common measures of educational outcomes. Disagreements about the appropriate objectives for public education, state fears of external evaluation, and objections to federal control all interfered with efforts to reach agreement on common metrics. Governors, long the leading opponents of national educational measurement, laudably took leadership in support of standardized outcome metrics, ultimately resulting in the compromise agreement of The No Child Left Behind law passed in 2000. While the law calls for states to set their own goals and develop their own systems to measure progress toward the goals, it also requires every state to participate in a national math and language-arts test for 4th- and 8th-graders.

The Education Trust, working with the American Institutes for Research, normalized data across states by looking at the performance of individual schools in each state relative to the performance of all schools in the state. This methodology enabled the identification of top-performing schools relative to other schools. Studying these schools relative to standardized school input variables and student background characteristics should help researchers identify common strategies successful schools use.

Actions to Implement Recommendation 3

- **Federal agencies** should aggressively encourage the generation of common state metrics.
- Although states are often seen as opponents of common measures, state leaders have on occasion in the past—and should with more regularity in the future—encourage information standardization to allow the research that will enable states to distinguish effective programs from ineffective ones. **State leadership** should come from the governors as well as program and information system managers.
- **National associations of state officials** should also play a leadership role promoting common performance metrics. NGOs should also try to fill in conceptual gaps providing federal agencies and states useful models for standardization and normalization approaches when the governments have not developed them.

Recommendation 4: Require Measurement

Federal agencies often find it difficult to carry out common measurements across states without a congressional mandate to do so. State leadership on data standardization can sometimes obviate the need for congressional action. In most cases, however, except where collection of state performance data is an old and established tradition as with FHWA, standardized state performance measurement is more likely to happen if Congress mandates

Actions to Implement Recommendation 4

- When common metrics do not already exist, **Congress** should mandate and financially support the full spectrum of state performance measurement in areas where federal agencies depend on states to accomplish their objectives, along with annual training of state information handlers.
- **States** should assume a leadership role, even in the absence of congressional mandates, to standardize their data and encourage federal agencies to collect, organize, and analyze it to help states identify effective practices worthy of replication and ineffective ones that should be curtailed.

its generation, collection, and dissemination. Not surprisingly, states more readily accept those mandates when they are accompanied by federal funds, although federal funds are not always essential.

For example, the National Highway Traffic Safety Administration directly funds and trains state staff charged with collecting state and local data about fatal accidents and coding the data into the federal system. To improve data comparability and quality, NHTSA provides and requires state employees to participate in a week of training each year. Congress mandates its collection and supports that effort by funding a staff position in each state for data collection.

Federal mandates to use measurements can also stimulate private vendors to create audience-focused, value-adding computerized performance management systems. Mandates for states to use the performance data to create transportation management systems made it possible for private sector consultants to mine federal collections of state data, then develop and sell management systems to help metropolitan areas address their transit planning needs.

Recommendation 5: Involve and Benefit Those Being Measured

Federal agencies often hesitate to gather and disseminate state performance information for fear of provoking state ire, and can be even more reluctant to use comparative performance information as the basis for decisions about federal intervention in state program design and implementation. Unfortunately, federal failure to gather, organize, and disseminate state performance information undermines the potential benefit of the information. It reduces the likelihood that anyone can learn from the experiments taking place in what Justice Louis Brandeis described as "the laboratories of democracy"—the states.

Fortunately, the federal government can take many actions to minimize the political volatility of state performance measurement systems without compromising its motivational or instructional value. One of the key tools is engaging those being measured in the design of the measurement system and the design of its analysis. This increases the value of the measurement system to the states, and builds their trust in its fairness.

Even when states have been engaged in efforts to develop common measurements or systems for managing with them, states may resist federal measurement efforts if they perceive it as an encroachment on areas traditionally managed solely by state and local agencies. In those instances, congressional mandates for common state measurement systems are extremely valuable. Even then, however, it is critical to work with the states to make the measurements useful to them lest they organize to reverse the congressional mandate. It is also critical to communicate the lessons of the measurement to the public, so they will benefit from and recognize the importance of collecting them.

NGO efforts to compare states may be greatly strengthened when they engage states in the development of the comparative metrics and when they include positive as well as negative findings.

For example, EPA jointly developed the NEPPS framework with states, leading more than half of the states to attempt a shift in their management focus from an emphasis on activities to an emphasis on environmental outcomes. As part of the implementation of that framework, EPA and the states reached agreement on a set of core performance measures. Separate from but related to this effort, they also agreed to establish a new organization, the Environmental Data Standards

Actions to Implement Recommendation 5

- **Federal agencies** should routinely engage states as co-owners of performance data and as decision makers in developing tools to enhance the use of the data, especially to serve state needs. When federal agencies lack authority to set national standards, they should encourage states to develop their own data and performance standards and support states' efforts to do so, while maintaining sufficient influence to assure the state-set performance standards continually improve social outcomes relative to existing standards and conditions. Engagement of the states should not preclude involvement of other parties who can also use the gathered information to advance outcomes. When supporting state development of data or performance standards, federal agencies should take care to consult Congress, public interest groups, and others whose actions affect outcomes or who depend on performance information to inform choices among options.
- **States** should seek support from the federal government to support state efforts to develop, adopt, and collect common measures if the federal government has not offered it. Support can come in the form of secretariat services to assist the states and funds to hire consultants to support state data and even performance standardization efforts.
- **Congress,** through funding and legislative language associated with funding, should support federal agency efforts to engage states in the selection and development of data standards and analytic tools.

Council, to agree on common definitions for state measurement efforts.

A nonprofit environmental advocacy organization, the Michigan Environmental Council, engaged former state officials in the development of common and normalized environmental performance metrics for states. Its report offers a promising model that EPA and the states can build upon in the future for creating a compendium of state information.

Recommendation 6: Encourage Analysis

Federal analysis of the full spectrum of state performance information adds great value to performance measurement, whether the federal government conducts the analyses itself or encourages others to do so. Analysis translates raw performance information into "actionable" information that motivates and illuminates. Good analysis sheds light on the effectiveness of interventions and helps those who must choose among options find the alternative best suited to their own tastes. It also helps identify areas with the most significant problems warranting greater attention. Good measurement and analysis builds support for more good measurement and analysis both in Congress and among "measures."

Useful analysis can take many forms. It need not be highly complex statistical analysis. Simpler forms of analysis, such as the identification of strong performers and weak performers, the relative frequency of problems, and the organization of performance data by the categories that might explain performance variations trigger useful follow-up questions that lead to valuable insights for improving performance. Comparisons across peers have motivational value if done in a manner that is perceived as fair (similar entities are being compared) to those being measured. More sophisticated statistical analyses (e.g., regression analysis and studies with control groups) can more precisely isolate variables likely to influence performance levels, and help states make more informed decisions to manage complex systems.

Federal agencies can find it politically difficult to do some types of analysis themselves because their findings get interpreted as judgmental or political. Funding external parties to conduct research can sometimes lessen these tensions without compromising the quality of the analysis. Federal agencies can further boost the value of and support for federal collection of state performance information by supporting state and local efforts to develop analytic tools that help them use the measures gathered to run their programs more effectively and efficiently.

For example, analysis by NHTSA and FHWA of the characteristics and consequences of fatalities, as well as state laws and practices that explain variations in fatality rates, makes it possible to identify the leading contributors to fatalities and serious injuries. With this information, FHWA and NHTSA have been able to promote national standards and

Actions to Implement Recommendation 6

- **Federal agencies** should analyze performance information to identify programmatic interventions that work, those that do not, and problem areas that need attention. Through its analysis, the federal government should seek to make performance information "actionable." That is, it should inform choices, answer questions of key decision makers, and trigger focused follow-up inquiries.
- Through ease of data access and funding for external research, **federal agencies** should encourage states and others to conduct similar analyses to find both the most effective and problematic practices. The federal government should experiment with funding a wide variety of parties to conduct external research to increase the prospects for audience-tailored analysis, incorporating quality-control mechanisms to enhance objectivity.
- **Non-governmental organizations** should analyze state performance data to find and commend state successes, in addition to identifying and highlighting states with poor performance. This may motivate in a less contentious manner, increasing prospects for greater state acceptance of NGO analysis.

programs that steadily reduce the nation's traffic fatality rate while improving its mobility.

Likewise, the Department of Education's wall chart with comparative state data survived through two Secretaries of Education. As so often happens when comparative measures have been generated without legislative mandate or when a well-organized audience fails to rise up in arms when measurement stops, those being measured and program managers in the department who saw little value from the chart succeeded in killing it. More recent department efforts to fund parties outside the agency to compile and analyze state performance data appear to be less controversial.

Recommendation 7: "Market" the Results

When federal agencies "market" measurements and what they learn from them in a way that successfully reaches those being measured and other key audiences, it increases the political prospects for a long-term federal/state performance measurement collaboration of the sort needed to contribute to continual performance gains. Key audiences that agencies should target include the federal and state legislators that authorize and appropriate funds for agency action. Effective marketing necessitates an understanding of the audiences that can contribute to better outcomes, the needs and interests of those audiences, and the information channels to reach them. Federal agencies have demonstrated the value of marketing a variety of different kinds of products to help states and localities improve outcomes, including compendia of state measurements, analyses of the measurements, systems to support analysis of the measurements to make them actionable, and programs that evidence suggests are effective. Federal agencies can promote the latter by packaging campaign kits for use by states and other key delivery partners, complete with explanatory evidence about the need for the program, training or instructions to run the program, public outreach

Actions to Implement Recommendation 7

- **Federal agencies** should make information dissemination, in an audience-focused format, a priority. They should make it a priority to return the information they gather to those who supply it (states, local governments, etc.) with value added, helping them learn from the experience of other states and even their own. They should also share performance information with Congress in a format that quickly answers questions Congress has. This may imply displaying performance information with greater geographic specificity so those in Congress can relate to it. It also implies delivering information in a timely manner, especially when the information is needed to inform policy debates.
- **Federal agencies** should work with states to explore if and how they can organize and analyze performance information to answer questions of state legislators and other key state officials.
- **Federal agencies** should also take care to analyze and disseminate information to serve the needs of federal agencies, including both their program and regional offices, so they will support a strong performance measurement system.

material states can use in their implementation efforts, and post-program assessments.

For example, NHTSA runs several outreach campaigns to help states increase seat belt use. It plays the role of wholesaler, providing states with marketing plans and materials. The "Buckle Up America" campaign, for example, provides states and others with well-developed marketing materials and plans to help them increase seat belt use. As part of that effort, NHTSA and states test and evaluate a carefully designed "Click It or Ticket" campaign that combines public relations, paid marketing, enforcement, belt use observation surveys, and public awareness surveys.

NHTSA regional administrators work closely with their states, sending memos that celebrate those with stronger performance while encouraging those with weaker performance to adopt new practices that have improved safety records in other locales.

Recommendation 8: Motivate with Comparison and Rewards, but Carefully

The display of comparative information can kindle the competitive spirit of states—motivating those that like to be pack leaders to strive for the top, encouraging friendly peer rivalries among similar states, and tapping into the fact that no state likes to be at the back of the pack. It can also inform individual decision making, creating a market-like mechanism when similar services are available from multiple governments or from multiple locations of the same government.

Careful competition necessitates fair comparisons and sensitive presentation of the information. Federal agencies may be able to diffuse some of the political tension surrounding interstate comparison without compromising its motivational and instructional potential by reporting in one place the full spectrum of state measurements, including information about various outcomes of public concern, outputs, activities, inputs, and relevant background characteristics. Creating a compendium of state measurement will show the strengths of state programs along with their weaknesses. Regional offices can motivate less contentious competition by conducting comparisons of smaller groups of states, keeping them "below the radar screen."

Rewards and sanctions linked to the attainment of a goal or linked to relative performance status can work. They must be used with care, however, lest the exercise of sanctions provoke so much resistance from those being measured that they work to undermine the system, either by organizing to dismantle it or by paying so little attention to the quality of the measurement they submit that it has little value. If there is insufficient public support for the goal to which the sanction is linked, an aggressive opponent on a key legislative committee, or a large number of states at risk of being sanctioned, the incentive system may be at risk. Agency efforts to market analyses showing how certain government programs are performing can improve support for the use of sanctions. Nothing, however, guarantees that every incentive system will work and will survive.

Actions to Implement Recommendation 8

- **Federal agencies** should hone their skills to provide balanced presentations of comparative performance information, including developing the skills and capacity of their regional offices to coach and encourage friendly competition among small groups of states. National program offices should support this effort through information organization and analysis, including the identification of relevant peer groups.
- **Congress** should give federal agencies the power to reward states for improving their social conditions and to sanction them for the failure to attain certain goals or adopt programs demonstrated to be more effective than what the states are doing, but also give the agencies authority to waive the sanctions when a state has made significant progress, even if it has not attained the goal or demonstrated significant effort to attain the goal.
- **Federal agencies** should use sanctions as a threat and last resort, but lead with efforts to identify and market effective programs.
- **Congress and federal agencies** should explore increased use of incentives structured similar to the seat belt program, which mandate comparable measurement and reward performance relative to peers and relative to past performance.

For example, international and private sector comparisons of student achievement initially and continually motivate government efforts to improve the American educational system. Although states initially asked the Department of Education to help them understand the causes of poor educational performance, many were unhappy when the agency released the comparative information to the public. The original department compendium of comparative information did not survive its originating sponsor's departure. Sustained parental concern about education quality, however, eventually created sufficient political support for a congressional mandate for a measurement system that will allow cross-state comparisons.

In contrast, Congress has occasionally been compelled to repeal goals it sets, including the goal that states establish statutes for a 55-mph speed limit and the use of helmets by all motorcycle riders. A lack of broad public support for the constraint on personal freedom most likely contributed to their repeal.

Congress was forced to repeal the mandate for six transportation management systems after states protested. The lack of public understanding of the value of the management systems may explain their rapid repeal. At the same time, the initial mandate successfully motivated most states to adopt the management systems for long enough that most of them began to appreciate their value, and have continued to use them.

Recommendation 9: Share Best Practices

This report has only begun to tap into the rich lessons of past experience pertaining to federal agency handling of state performance goals and measures to enrich public well-being. This report suggests that the federal government can play a powerful role supporting learning across governments to speed the identification and adoption of effective programs. It is reasonable to believe that, similar to the way states can benefit from the federal government playing an information gathering and brokering role, federal agencies would benefit from an organization gathering and brokering learning across federal agencies.

The federal government has a choice. It can allow each agency to experiment independently in its own work with states and localities using performance goals and measures to improve outcomes. Or, it can try to facilitate communication and evaluation across federal agencies to speed movement along the learning curve. The National Academy of Public Administration and the General Accounting Office have already begun noteworthy efforts to document and disseminate federal agency experience implementing GPRA, and have directed some attention to the issue discussed in this report—how federal agencies could use performance goals and measures in their work with states and localities to improve societal outcomes. Given how much federal domestic policy agencies depend on other levels of government to accomplish their objectives, the general question of effective methods for federal handling of state and local performance goals and measures cries for more attention.

The Office of Management and Budget (OMB) has issued guidelines for federal agencies specifying how they should handle GPRA and has graded agency performance relative to specific programs. It has not, however, directed much attention to the powerful role it could play facilitating learning across agencies.

OMB could more fully try to understand agency experiences such as those discussed in this study. For example, building on the analysis started here, it could track the ways federal agencies have historically structured the relationship between goal attainment and incentives (both rewards and penalties), the structures that have worked (motivated outcome improvements) and those that have not, and the characteristics of each situation that affected the results. Then, as agencies and Congress consider new federal policies using goals, performance measures, and incentives, they could turn to OMB for expertise about options likely to be effective and sustainable.

Taking on this responsibility will not be easy for OMB; its staff tend to operate along agency lines, rather than across agencies. OMB staff also tend to structure most decisions in the context of budget transactions. Nonetheless, OMB is the logical agency within the executive branch to look across agencies and accumulate expertise in this area, filling a badly needed function. One way OMB might transition to this kind of work would be through

Actions to Implement Recommendation 9

- **Congress** should mandate and fund a learning capacity across the federal government to identify and disseminate effective ways federal agencies can use performance goals and measures in working with states and local governments.
- The **Office of Management and Budget, the National Academy of Public Administration, and the General Accounting Office** should each contribute actively to identify effective ways federal agencies can use performance goals and measures in their work with other levels of government, marketing what works so other agencies can replicate it, and detecting what does not so ineffective approaches are not repeated.

sponsored research, coupled with training for its staff about the research findings. Given the potential for performance goals and measures to improve societal outcomes, the emphasis the George W. Bush administration has placed on performance management, and technology advances that have dramatically improved the economics of analyzing and disseminating performance information, it is timely for OMB to assume a stronger role facilitating analysis and learning across agencies regarding effective ways to use state performance goals and measures.

Federalism is not going away. Happily, performance goals and measures seem to be taking hold. It is therefore time to strengthen federal agency capacity to understand how to manage performance effectively in a federalist system. It is hoped that this report will catalyze conversation in Congress and federal agencies, among the states, and in NGOs about the question, "What and how can the federal government best use performance goals and measures to work with states to improve societal outcomes?" It is also hoped that, beyond debate, the ideas will inspire readers to act and tap the power of performance measurement and the dynamic tensions of the federal/state partnership to deliver better outcomes to the American people.

Appendix I: Summary of Implementation Actions by Key Actor

Congress

- Congress should require federal agencies that work with states to publish annual (or biannual) compendia of state-specific performance information. It should provide funding and training to state staff for that purpose.
- When common metrics do not already exist, Congress should mandate and financially support the full spectrum of state performance measurement in areas where federal agencies depend on states to accomplish their objectives, along with annual training of state information handlers.
- Congress, through funding and legislative language associated with funding, should support federal agency efforts to engage states in the selection and development of data standards and analytic tools.
- Congress should give federal agencies the power to reward a state for improving social conditions and sanction it for failure to attain a goal or adopt a program demonstrated to be more effective than what a state would otherwise have done. It should also give the agencies authority to waive sanctions when a state has made significant progress or demonstrated a significant effort to attain a goal, even if it has not attained it.
- Congress should explore increased use of incentives structured similar to the seat belt program, which mandate comparable measurement and reward performance relative to peers and relative to past performance.

Federal Agencies

- Federal agencies should make it a top priority to collect and organize state performance information in a readily accessible, easy-to-use data repository.
- Each federal agency should make clear in its annual GPRA performance report what relevant, state-specific performance information is available and where to find it.
- Federal agencies should include in their data repositories the full spectrum of information from states—about outcomes, intermediate outcomes, outputs, and the inputs of state programs, as well as the characteristics of the people, places, and things being affected by or potentially affecting program outcomes.
- In some cases, federal agencies may want to fund third parties to create the state databases.
- Where federal agencies lack the full spectrum of state measurement data, they should take the necessary actions to supplement them, including seeking congressional support and mandates.
- Federal agencies should aggressively encourage the generation of common state metrics.
- Federal agencies should routinely engage states as co-owners of performance data and as decision makers in developing tools to enhance the use of the data, especially to serve state needs.
- Federal agencies should encourage states and others to conduct similar analyses to find both the most effective and problematic practices.

- The federal government should experiment with funding a wide variety of parties to conduct external research to increase the prospects for audience-tailored analysis, incorporating quality-control mechanisms to enhance objectivity.
- Federal agencies should analyze performance information to identify programmatic interventions that work, those that do not, and problem areas that need attention.
- Federal agencies should analyze information to serve their own internal needs, including the needs of both headquarters and regional offices, so all will contribute to a strong performance measurement system.
- Federal agencies should make information dissemination, in an audience-focused format, a priority.
- Federal agencies should work with states to explore if and how they can organize and analyze performance information to answer questions of state legislators and other key state officials.
- Federal agencies should hone their skills to provide balanced presentations of comparative performance information, including developing the skills and capacity of their regional offices to coach and encourage friendly competition among small groups of states.
- Federal agencies should explore increased use of incentives structured similar to the seat belt program, which mandate comparable measurement and reward performance relative to peers and relative to past performance.

Cross-Government Entities

- The Office of Management and Budget and the General Accounting Office should facilitate learning across the federal government on the best ways to handle state information to advance outcomes.
- The Office of Management and Budget, the National Academy of Public Administration, and the General Accounting Office should each contribute actively to identify effective ways federal agencies can use performance goals and measures in their work with other levels of government, marketing what works so other agencies can replicate it and detecting what does not so ineffective approaches are not repeated.

States, State Associations, and Non-governmental Organizations

- State associations should support and work with federal agencies in the effort to collect and analyze useful state performance and other information.
- State leaders and their associations should work with the agencies to craft workable proposals for congressional mandates to create useful compendia of state performance information. States should also work with federal agencies to build repositories of state data even in the absence of mandates.
- National associations of state officials should also play a leadership role promoting common performance metrics. NGOs should also try to fill in conceptual gaps, providing federal agencies and states useful models for standardization and normalization approaches when the governments have not developed them.
- State leaders—governors as well as program and information system managers—should encourage information standardization to allow the research that will enable states to distinguish effective programs from ineffective ones.
- States should assume a leadership role, even in the absence of congressional mandates, to standardize their data and encourage federal agencies to collect, organize, and analyze it to help states identify effective practices worthy of replication and ineffective ones that should be curtailed.
- States should seek support from the federal government to support state efforts to develop, adopt, and collect common measures if the federal government has not offered it.
- Non-governmental organizations should analyze state performance data to find and commend state successes, in addition to identifying and highlighting states with poor performance.

Appendix II: Synopsis of Eight Key Components of EPA's National Environmental Performance Partnership System

The National Environmental Performance Partnership System (NEPPS), as adopted, had seven principal components and a complementary grant mechanism. What follows is a synopsis. For a copy of the full policy, see www.epa.gov/ocir/nepps:

1. *Increased use of environmental goals and indicators.* EPA and the states would develop, jointly, common performance measures each state would report for purposes of national environmental assessments. In addition, each state would identify appropriate state-specific environmental performance indicators. The indicators, which could change over time as experience was gained, would serve as the basis for making decisions about state and EPA activities in each state each year. The measures would be regularly collected and made available to the public. The states would work with EPA in an equal partnership to select, test, develop, and adopt the indicators and measures.

2. *New approach to program assessments by states.* States would begin to conduct self-assessments of their own progress, and share them with the public. NEPPS called for self-assessments that would include an assessment by each state of its key environmental problems and opportunities; a description of the recent performance of the state's programs based on available measures of program success including an analysis of current program weaknesses; an assessment of basic fiscal accountability; and the state's proposed action plan for maintaining and improving its environmental program performance, identifying specific actions and approaches planned for the coming year; suggestions for ways EPA could assist the state in improving performance or achieving stated goals; and a report on how well the state carried out the prior year's Environmental Performance Partnership Agreement. In addition, the NEPPS agreement called for the states and EPA to explore the use of visiting program evaluation teams to each state composed of both state and regional staff.

3. *Environmental performance agreements.* Based on environmental conditions, the state's self-assessment, and EPA's assessment, each state and EPA would sign an agreement regarding appropriate national and state-specific environmental goals, program-specific and multi-media performance indicators, state commitments for specific deliverables and activities to address identified needs, the allocation of federal resources to shared goals and priorities, disinvestments, and commitments for federal assistance. The agreement would take precedence over the existing program work plan process.

4. *Differential oversight.* EPA's program evaluation would shift from a review of specific permits, inspections, and enforcement actions states had already taken to after-the-fact, program-wide assessments. Based on those reviews, EPA would exercise differential oversight, treating states differently based on their prior performance. Where the reviews indicated poor state performance, EPA would return to more case-

specific reviews and interventions. For those performing well, it would continue to conduct after-the-fact reviews to confirm continued strong performance, taking direct implementation actions in states only upon request of the state. Some exceptions were noted, such as when a firm operating in several states showed problems that could best be addressed with a national enforcement action.

5. *Performance leadership programs.* In addition to differential oversight, the NEPPS system also called for a label to recognize publicly as "leaders" state programs that were historically strong. Those earning the label would not need routine federal oversight. States would apply for the designation. Also, states and EPA would work together to define the criteria for earning the leadership label.

6. *Public outreach and involvement.* EPA and the states committed to discussing the NEPPS system, as well as individual state self-assessments and the priorities reflected in the annual state/EPA performance agreements, with the public.

7. *Joint system evaluation*. As the new system is implemented, EPA and the states agreed to review the results and experiences to ensure continuous improvement. Criteria for assessment would include effectiveness, public credibility, and fiscal soundness.

8. *Performance Partnership Grants.* Concurrent with and in support of NEPPS, EPA sought and obtained congressional authorization giving states the ability to opt to combine federal grant money from 16 categorical EPA grants into a single grant, essentially creating a state-triggered, discretionary block grant.

Appendix III: The Early Federal Road Organizations: A Cooperative, Information-Rich Organizational Culture[65]

Performance measurement is not a new business for the federal government, nor is the need to address how the federal government uses performance measures to work with the states. To appreciate fully how the federal government can fruitfully use performance goals and measures in its work with states to improve societal outcomes, lessons from earlier periods of American history are also instructive. Examples from over a century ago—when the first federal road agency, the Office of Road Inquiry (ORI), was established—richly demonstrate that when the federal government gathers performance information that helps states and Congress understand which intervention strategies work and which do not, both states and Congress are more likely to value the federal role in performance measurement and support it.

The Office of Road Inquiry was established in 1893 in the U.S Department of Agriculture. Somewhat ironically, given today's perceptions of the priorities of the Federal Highway Administration, the original advocates for the federal highway department were bicycle riders seeking government funds to improve local roads in rural areas for recreational enjoyment. The cyclists hoped to build an alliance with farmers, whom they thought would also value improved rural roads. The cyclists and their main advocacy group, the League of American Wheelman, initially failed to convince the farmers and state legislatures of the need for funding. By 1893, however, they had convinced Congress to appropriate $10,000 for the Department of Agriculture to investigate road construction and management around the country. The 1893 Agricultural Appropriation Act instructed the Secretary of Agriculture:

> … to make investigation in regard to the best method of road-making … and to enable [the Secretary] to assist agricultural college and experiment stations in disseminating information on this subject.…[66]

With its first appropriation for federal involvement in roads, Congress made clear its intent that the Secretary of Agriculture was to model the federal road effort on the agricultural land-grant experience. There, the federal government worked closely with state and local institutions to gather performance information about various agricultural practices, document and evaluate them, encourage experiments in better performing practices, and broadly disseminate best practice information through field applications that could easily be seen and replicated as well as through written documentation.

A Wheelman activist, Roy Stone, was named the first head of ORI. At the time ORI was established, only two states had road departments.

Information Gathering and Dissemination

The Secretary of Agriculture gave Stone, a civil engineer, clear instructions: gather information and avoid politics. With only a clerk and himself as staff and virtually no budget, Stone reached out to harvest all the relevant information he could find about the nation's rural roads. With few counterpart state road

agencies, Stone sought information from governors, secretaries of state, congressmen, state geologists, and railroad presidents. He sought information on state road laws, the location of road materials suitable for building, and rates for rail haulage.[67] Within a few years, ORI was collecting county road maps as well. When ORI could not gather data about factors contributing to road performance, such as the traction performance of different road surfaces, it would conduct its own field tests.

ORI quickly compiled and organized the data it gathered into bulletins and circulars; 18 of the former and 23 of the latter were issued within its first two years of operation. By gathering, organizing, and disseminating information no one else had attempted to gather, ORI provided valuable information to Congress, industry, and the states, establishing itself as both an expert and valued resource.

To share what it had learned about factors contributing to road performance with the broader public, ORI began working with local communities in 1896 to construct quarter-mile "object-lesson" roads using the best-known construction methods. ORI provided supervision and technical assistance if a community supplied the labor and materials. By 1899, 21 object-lesson roads had been built in nine states. The program proved extremely popular; dozens of other communities sought to create similar working partnerships with the federal government when their resources allowed.

In 1904, the Office of Road Inquiry conducted the first complete inventory of rural roads in the United States. To do that, the federal roads office reached out to local governments for assistance and cooperation. It contacted every county official in the country, seeking information on mileage, road surface type, revenue sources, expenditures, and levels.

Despite the admonition to avoid politics, neither Stone nor his successor remained altogether apolitical. They maintained informal alliances with the League of American Wheelman and other activists and built alliances with newly created state road agencies, creating a powerful coalition able to influence federal legislation and win federal funding for many years.

Information Analysis and Experiments

Logan Page, who headed the ORI's successor agency between 1905 and 1918, strengthened the federal road office's role as an expert resource for state and local governments. He continued the federal office's data collection efforts and its dissemination of written materials on best practices. He also continued to provide federal experts to state and local governments willing to provide materials and labor costs to build local object-lesson roads. In addition, he greatly enhanced the office's capacity to assess road quality. Under Page, the federal road agency expanded its testing laboratory, developed equipment and methods for testing road quality, and conducted field-based road performance investigations. Page further enhanced the role of the federal roads office as "the bully pulpit of expertise."

Best Practice Guidelines

The increased demand for the road-building instructions offered since 1895 eventually led the federal road agency to identify state-of-the-art models and recommend minimum performance conditions that state and local governments (and others) might want to incorporate into their own purchasing requirements. The federal government produced its first set of voluntary standards for road materials and testing procedures in 1911. Voluntary construction guidelines and bridge specifications soon followed. Although other industry organizations and a few states had also developed their own voluntary standards, the knowledge accumulated by the federal office through its testing procedures gave its standards greater credibility. They were often treated as the accepted standard and even promoted by professional and trade associations. States with their own road offices quickly adopted standards similar to those suggested by the federal government.

Procedural Standards to Improve State Capacity

In an effort to improve the quality of the roads, the federal roads office also developed standards suggesting minimum characteristics needed for an adequate state highway office, including necessary

skill training for staff. Under Page's leadership, the federal roads office promoted the creation of state road departments to provide engineering expertise and urge state aid to local governments. By the first decade of the 1900s, only a handful of states had established roads offices and virtually no states had funded road building or even technical assistance programs. Page saw this as a serious impediment to achieving better quality roads. To address the problem, his office developed and promoted model legislation for states to create road-aid programs and state road agencies. Many states used the model.

Federal Road Aid Act of 1916

After the turn of the century, automobile owners began pressing states and the federal government for better roads in the rural regions surrounding and connecting urban areas. In 1911 and 1912, 60 road bills were introduced in Congress. Congress passed the first federal funding for roads in 1913, when it appropriated $500,000 for a demonstration program. By 1916, Congress expanded its funding commitment with the Federal Road Aid Act of 1916. The act established the basic structure of the federal-aid highway program that, with some modifications, still operates today.

Many, including the American Automobile Association, sought direct federal funding for roads. The federal road agency, however, did not want to run a program placing it directly in charge of road construction, or even working directly with local governments to guide their road-building efforts. From past experience, it was concerned about the difficulty of managing such a massive field operation. The federal road agency had already begun to work constructively with the state road agency directors that did exist, so its leaders pressed for a federal law that worked through and strengthened the capacity of the states. Congress adopted the law the federal road agency sought, giving the states responsibility to oversee the construction of roads in their state. The law provided up to $10,000 per mile in federal funds for rural roads as a match to state expenditures. Funds would be allocated to states using a formula based on state characteristics, but only after the state had established a highway department that met minimum federal standards.

Mandates to Improve State Capacity

With increased federal funding came increased federal control over the use of the funds. The federal government reserved the right to inspect all state-approved roads and conducted work to assure it met federal standards. It also approved the location and construction methods for federal-aid roads, refusing funding for those that failed to meet federal expectations.

Shared Authority/Shared Obligations

Page's successor, Thomas H. MacDonald, established another very important pattern still evident in the way the federal highway agency currently uses performance measurement in its work with states: collaborative decision making. His long tenure at the federal highway agency, from 1919 until 1953, assured that MacDonald's commitment to federal/state collaboration took firm hold.

Prior to arriving in Washington, MacDonald had managed the Iowa state road agency, which at the time was considered among the most sophisticated of state road offices. As a state leader, MacDonald had worked closely with Page to shape federal highway law. He understood the potential value of state support for federal efforts in Washington. As a state official, he had also seen how the increased powers given the federal government under the new law had heightened federal/state tensions. He chose to tackle this emerging tension by engaging states more directly in federal decision making and promoting a more helpful federal attitude toward the states.

As a condition of taking the job in the federal highway agency, MacDonald insisted on the creation of a Federal-Aid Advisory Committee composed of six state highway engineers. This early overture to the states was not mere window dressing, but rather intentional pattern setting. Several subsequent decisions—sharing standard-setting responsibility with the states, insistence on cooperative rather than controlling language by federal officials, and collaboration with states in the conduct of research—illustrate how MacDonald promoted a full federal partnership with states focused on improved results.

As a state official, MacDonald had long advocated the creation of a state-run standard-setting commit-

tee. In 1919, after assuming the helm of the federal agency, MacDonald asked the American Association of State Highway Officials (AASHO) to form a permanent standards committee to help the federal government update standards for roads, materials, and bridges under the new federal highway law. AASHO accepted.

AASHO reciprocated the federal agency's overture to the states with its own overture to federal officials for assistance and expertise. It turned to senior federal officials to chair its technical committees since few state legislatures were willing to fund state staff to work on projects that served other states. MacDonald himself chaired the AASHO technical standards committee its first three years. The federal agency provided the secretariat staff for the AASHO committees, as well. Further, federal staff provided the background data used by the technical committees, including fleshed out proposals for discussion.

Even with significant federal assistance and guidance, the states retained decision-making authority in proposing the standards since states comprised the committee membership that actually decided the standards. Once the standards were established, the federal agency still needed to follow federal rule-setting procedures to determine whether to adopt the state-recommended standards as its own. The structure of the relationship assured both parties strong influence. And, AASHO's role in developing the standards reduced the states' inclination to blame the federal government for decisions that would undoubtedly impose unwelcome changes on some of the states.

MacDonald insisted on state participation in all aspects of federal decision making. Following the adoption of the Federal Highway Act of 1921, staff of the Bureau of Public Roads (BPR) proposed revisions to the federal aid regulations and submitted them to MacDonald. After reviewing the proposed revisions himself, MacDonald submitted them to the AASHO executive committee to "discuss in detail." The final version of the regulations, which originated in the federal agency, incorporated the changes AASHO had suggested. This review procedure became standard practice, sending a strong signal throughout the federal road agency that state officials needed to be consulted prior to making significant policy changes.[68]

Cooperative Learning Capacity

MacDonald also worked hard to include states in the conduct of roads research. Under Page, the federal road agency had already built a collaborative research alliance with the highway construction industry. MacDonald sought to bring states into the learning system. After a number of road failures that proved embarrassing, states clamored for more research. The agency created a research magazine, *Public Roads,* to share its knowledge more systematically. In addition, it decided to strengthen highway research.

Again, MacDonald sought to establish a cooperative federal/state venture, both to assure that the direction of research would meet state needs and to buffer the federal agency against state attack. Historian Bruce Seely writes:

> His reticence about BPR coordination of a program of highway research did not stem from doubts about the value of research. Rather, he knew that a coordinator might have to set tasks that would inevitably attract complaints from unhappy state engineers and other researchers. Therefore, he declined to establish the bureau as the official highway research center. Instead, he … acted as midwife for a new agency within the National Research Council that mirrored the cooperative principles of federal aid….
>
> From the start, MacDonald envisioned a highway research group in the associative mold, functioning as a clearinghouse for research results rather than a rigid, centralized coordinator.[69]

The official history of the Highway Research Board describes the benefits of this cooperative structure:

> Decentralization could enlist the interest and aid of more people, it could encourage more individual initiative, it could provide for the important investigation of peculiar local problems, it could institute cooperative research activities in its own locale, it could spread efforts necessary on nationwide problems, it could profit by blending viewpoints from many regions

and institutes, and finally it could and would use the funding from its own efforts in its own operations.[70]

Cooperation without Compromising Control

MacDonald's decisions to share power and problems with the states did not mean he abdicated the federal ability to assess performance and motivate improvements. The federal road agency continued to monitor state road conditions and deploy federal engineers to inspect roads. When federal inspectors saw problems, they followed up and, if necessary, withheld payments of federal funds. Under MacDonald, the agency continued to treat its audit and oversight responsibilities seriously.

But under MacDonald's leadership, the federal road agency made a subtle but significant change in the tone of the federal approach to the states: "Whereas [the federal road agency under Page] had *demanded* changes in Michigan's specifications, the Bureau of Public Roads *requested* alterations."[71] Follow-up investigations were seen as a complement to the agency's primary emphasis on helping the states learn how to improve their performance by gathering information, conducting research, establishing standards of best practice, and promoting that knowledge.

In sum, throughout its first 50 years of operation, the federal road office made performance information the center of its strategy for working with states to improve the quality of America's roads. It annually gathered information from local and later state governments, the laboratory, and field experiments about road performance, characteristics of the roads, and local government road programs. It emphasized the information value of performance measures to create a sustainable culture of learning about and improving road performance. Historian Bruce Seely credits the first head of ORI, Roy Stone, with "pioneering three enduring patterns of activity … : Build[ing] a reputation for technical *knowledge*, *Promot[ing]* the gospel of good roads, and Utiliz[ing] *cooperation* to reach these goals."[72]

Endnotes

1. See, for example, Beryl Radin, "Intergovernmental Relationships and the Federal Performance Movement," *Publius: The Journal of Federalism* 30 (2000); Teresa Curristine, "Reforming the U.S. Department of Transportation: Challenges and Opportunities for the Government Performance and Results Act for Federal-State Relations," *Publius: The Journal of Federalism* 32 (Winter 2002); Sharon L. Caudle, "GPRA Relations: Suggested Federal-State Practices" (paper presented at the American Association of Public Policy and Management, Washington, D.C.: November 1, 2001).

2. David G. Frederickson, "The Potential of the Government Performance and Results Act as a Tool to Manage Third-Party Government," (Arlington, Va.: The IBM Center for The Business of Government, August 2001), p. 21.

3. The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (P.L. 104-193).

4. *Performance management* refers to the use of performance goals and measures for management. *Performance measures* refers to the full spectrum of information about outcome, output, activity, input, and relevant background characteristics.

5. Carol M. Browner, et al., "Joint Commitment to Reform Oversight and Create a National Environmental Performance Partnership System," policy memorandum signed at the EPA "All-States" meeting by Carol M. Browner, Administrator, U.S. Environmental Protection Agency; Tom Looby, Director, Office of Environment of the Colorado Department of Health and Co-chair of the State/EPA Capacity Steering Committee; Fred Hansen, Deputy Administrator, U.S. Environmental Protection Agency; and Mary Gade, Director, Illinois Environmental Protection Agency and Co-chair of the State/EPA Capacity Steering Committee; May 17, 1995. http://www.epa.gov/ocirpage/nepps/ovrsight.htm. [December 22, 2002].

6. Different federal statutes call for slightly different federal actions with regard to state implementation of the laws, including delegation, authorization, and granting of primacy. The terms are used interchangeably here, although they have slightly different meanings. EPA may grant full or partial delegation for each federal environmental statute.

7. Environmental Law Institute, "Comparison of Federal-State Allocation of Responsibility in Five Environmental Statutes" (Washington, D.C.: Environmental Law Institute, September 1995).

8. For a state view on the importance of federal oversight, see Mark Coleman, "EPA Oversight—The Baby and the Bathwater," *ECOStates* (Washington, D.C.: Environmental Council of States, Spring 2001), pp. 32-33.

9. Shelley H. Metzenbaum, "Making Measurement Matter: The Challenge and Promise of Building a Performance-Focused Environmental Protection System," Brookings Center for Public Management (CPM98-2), October 1998. Massachusetts had conducted a Ford Foundation/Kennedy School award-winning experiment, known as the Blackstone Project, to test the effectiveness of three different approaches to inspection and proposed the inspection approach its field research suggested would be most effective.

10. Cliff Rechtschaffen and David Markell, *Reinventing Environmental Enforcement and the State/Federal Relationship* (Washington, D.C.: Environmental Law Institute, 2003).

11. The Environmental Council of States, "Report to Congress: State Environmental Agency Contributions to Enforcement and Compliance" (Washington, D.C.: ECOS, April 2001), pp. 53-70.

12. http://www.epa.gov/edsc/

13. http://www.epa.gov/oei/imwg/files/Final_Blueprint.pdf

14. See Blaine Liner et al., *Making Results-Based State Government Work* (Washington, D.C.: The Urban Institute, 2001) for discussions of statewide trends in performance measurement. See Mia Seeley, "Compendium of State Environmental Performance Reports" (College Park, Md.: University of Maryland School of Public Affairs Environmental Compliance Consortium, draft 2002) for a list of state environmental performance reports. This report is currently being updated.

15. http://www.epa.state.il.us/environmental-conditions/2000/index.html

16. Joyce N. Ritter, "The History of Highway Statistics," *1994 Highway Statistics: 50th Anniversary Edition* (Washington, D.C.: U. S. Department of Transportation, 1994).

17. Ritter.

18. *"Public Roads*: 80 Years Old, But the Best Is Yet to Come, The Early History of *Public Roads," Public Roads* (May/June 1998, Vol. 61, No.6.). Available at: http://www.tfhrc.gov/pubrds/may98/histry.htm [Accessed October 17, 2002.] The quote is from Logan Page, the director of the office.

19. The preface to *Highway Statistics* for the year 2000 describes itself as the "56th of an annual series, it presents the 2000 analyzed statistics of general interest on motor fuel, motor vehicles, driver licensing, highway-user taxation, State highway finance, highway mileage, and Federal aid for highways; and 1999 highway finance data for municipalities, counties, townships, and other units of local government. Starting with the 1992 edition, International data has also been included." http://www.fhwa.dot.gov/ohim/hs00/preface.htm [accessed October 14, 2002.]

20. The "Status of the Nation's Highways, Bridges, and Transit: Conditions and Performance" report is a series of combined biennial documents prepared by the Department of Transportation that satisfy requirements for reports to Congress on the condition, performance, and future capital investment requirements of the nation's highway and transit systems. This report incorporates highway and bridge information required by Section 502(g) of Title 23 United States Code (U.S.C.), as well as transit system information required in 2000 by Section 308(e) of Title 49 U.S.C. The 1999 edition also includes the results of a study on Interstate Needs required by Section 1107(c) of the Transportation Equity Act for the 21st Century (TEA-21).

21. "Editor's Notes," *Public Roads,* November/December 2000 at http://www.tfhrc.gov/pubrds/nov00/endnotes.htm [Accessed October 17, 2002.]

22. See David T. Hartgen, "TEA-21 at Midpoint: Comparative Performance of State Highway Systems, 1984-2000," 11th Annual Report, 2002 (available from author, Professor of Transportation Studies, University of North Carolina at Charlotte, February 6, 2002, and at www.geoearth.uncc.edu/faculty/hartgen/TEA21atM.htm [accessed October 17, 2002]).

23. See, for example, "Important Information" in *Highway Statistics 2000* at http://www.fhwa.dot.gov/ohim/hs00/index.htm and http://www.fhwa.dot.gov/ohim/hs00/ps1.htm [accessed October 17, 2002.]

24. Federal Highway Administration, *Fiscal Year 2003 Performance Plan* (Washington, D.C.: September 2002).

25. U.S. General Accounting Office, "Transportation Infrastructure: States' Implementation of Transportation Management Systems" (Washington, D.C.: GAO/RCED-97-32, January 1997).

26. AASHTOWare, "History, Accomplishments," available at www.aashtoware.org/aashtoware/site.nst/all-pages/Accomplishments?OpenDocument [downloaded July 20, 2002].

27. AASTHOWare, "Overview," www.aashtoware.org/aashtoware/products.products.nst.allpages/overview [downloaded July 20, 2002].

28. U.S. General Accounting Office, "Transportation Infrastructure," p. 1.

29. The National Highway System Designation Act of 1995.

30. U.S. General Accounting Office, "Transportation Infrastructure: States' Implementation of Transportation Management Systems" (Washington, D.C.: GAO/RCED-97-32, January 1997).

31. GAO, "Transportation Infrastructure," 15-16.

32. *Standard setting* can refer to several different concepts: common metrics or data standards for measurement, minimum programmatic characteristics that must be fulfilled, or minimum performance outcomes that must be achieved. For an excellent discussion of federal standard-setting approaches vís-a-vís states, see U.S General Accounting Office, "Regulatory Programs: Balancing Federal and State Responsibilities for Standard Setting and Implementation" (Washington, D.C.: GAO, March 2002).

33. P.L. 89-564, section 402 (a). (U.S. Code Title 23)

34. P.L. 89-564.

35. National Highway Traffic Safety Administration, "Highway Safety Assessment: A Summary of Findings in Ten States," NHTSA Technical Report DOT HS 808 796 (Springfield, Va.: National Technical Information Service, June 1998), p. 5.

36. U.S. Department of Transportation, "National Highway Traffic Traffic Safety Facts 2000: Occupant Protection," (Washington, D.C.: U.S.DOT HS 809 327) available at http://www-fars.nhtsa.dot.gov/pubs/6.pdf [accessed November 10, 2002].

37. D. F. Preusser et al., "Evaluation of Motor Cycle Helmet Law Repeal in Arkansas and Texas" (Springfield, Va.: National Technical Information Service, September 2000).

38. By 1996, 25 states and the District of Columbia had adopted universal helmet laws, while another 22 had laws for younger riders.

39. D. F. Preusser et al.

40. Section 1403 of TEA 21 (Sec 157 of Chapter 1 of Title 23). See Fact Sheet for "Safety Incentive Grants for Use of Seat Belts" at http://www.nhtsa.dot.gov/nhtsa/whatsup/tea21/tea21programs/Factshee.157.html

41. www.nhtsa.dot.gov

42. http://www.nhtsa.dot.gov/people/injury/airbags/clickit_ticket/current.html

43. For a further discussion of the value of circumscribed competitions, see Robert H. Frank, *Choosing the Right Pond: Human Behavior and the Quest for Status* (New York: Oxford University Press: 1985).

44. See, for example, http://www.fhwa.dot.gov/ohim/hs99/index.htm [accessed July 21, 2002].

45. Janet A. Weiss and Judith E. Gruber, "The Managed Irrelevance of Federal Education Statistics," in William Alonso and Paul Starr, eds., *The Politics of Numbers* (New York: Russell Sage Foundation, 1987), pp. 363-391.

46. P.L. 107-110.

47. P.L. 103-382.

48. U.S. General Accounting Office, "Title I: Education Needs to Monitor States' Scoring of Assessments" (Washington, D.C.: GAO-02-393, April 1, 2002).

49. Weiss and Gruber, pp. 363-391.

50. Weiss and Gruber.

51. Weiss and Gruber (pp. 374-376).

52. Unless otherwise noted, the discussion in this section is based on Alan I. Ginsburg, et al., "Lessons from the Wall Chart," *Educational Evaluation and Policy Analysis* (Spring 1988, Vol. 10, No.1, 1-12.) See also, David Tyack and Larry Cuban, *Tinkering Toward Utopia: A Century of Public School Reform* (Cambridge: Harvard University Press, 1995), pp. 33-37.

53. SATs and ACTs are privately run tests students take as part of the college admissions process.

54. William T. Gormley, Jr., and David L. Weimar, *Organizational Report Cards* (Cambridge: Harvard University Press, 1999), p. 39.

55. National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform* (Washington, D.C.: GPO, 1983) available at http://www.ed.gov/pubs/NatAtRisk/.

56. Craig D. Jerald, *Dispelling the Myth Revisited: Preliminary Findings from a Nationwide Analysis of "High-Flying" Schools* (Washington, D.C.: The Education Trust, 2001), pp. 3-4.

57. Gormley and Weimar.

58. Craig D. Jerald.

59. The Education Trust, with cooperation of the Council of Chief State School Officers, "Dispelling the Myth" (Washington, D.C.: The Education Trust, 1991.)

60. Don McLaughlin of the American Institutes for Research, "State Assessment Data and Policy Research," Presentation at the National Center for Education Statistics Summer Data Conference 2002, Washington, D.C., July 25, 2002, http://208.253.216.16/Assessment/.

61. http://204.176.179.36/dc/edtrust/edstart.cfm [accessed November 15, 2002.]

62. U.S. Department of Education, Office of the Under Secretary and Office of Elementary and Secondary Education, "School Improvement Report: Executive Order on Actions for Turning Around Low-Performing Schools" (Washington, D.C.: January 2001).

63. U.S. Department of Education, "School Improvement…" p. 27.

64. http://www.nochildleftbehind.gov/media/news/070102.html [accessed November 15, 2002].

65. The information in this section is from Bruce E. Seely, *Building the American Highway System: Engineers as Policy Makers* (Philadelphia: Temple University Press, 1987).

66. Cited in Joyce N. Ritter, "The History of Highway Statistics," *1994 Highway Statistics: 50th Anniversary Edition* (Washington, D.C.: U. S. Department of Transportation, 1994).

67. Seely, p. 13.

68. Seely, p. 77.

69. Seely, p. 110.

70. Highway Research Board, *Ideas and Actions: A History of the Highway Research Board 1920-1970* (Washington, D.C.: 1970), p. 11. Cited in Seely, pp. 110-111.

71. Seely, p. 56.

72. Seely, p. 13.

Bibliography

Agranoff, Robert, and Michael McGuire, "American Federalism and the Search for Models of Management," *Public Administration Review* (November/December 2001), Vol. 61, No.6, pp. 671-681.

Caudle, Sharon L., "GPRA Relations: Suggested Federal/State Practices," paper presented at the American Association of Public Policy and Management, November 1, 2001.

Coleman, Mark, "EPA Oversight—The Baby and the Bathwater," *ECOStates* (Washington, D.C.: Environmental Council of States, Spring 2001), pp. 32-33.

Curristine, Teresa, "Reforming the U.S Department of Transportation: Challenges and Opportunities of the Government Performance and Results Act for Federal-State Relations," *Publius: The Journal of Federalism* 32 (Winter 2002).

Dye, Thomas R., *American Federalism: Competition Among Governments* (Lexington, Mass.: Lexington Books, 1990).

Education Trust with cooperation of the Council of Chief State School Officers, "Dispelling the Myth" (Washington, D.C.: The Education Trust, 1991).

Environmental Law Institute, "Comparison of Federal-State Allocation of Responsibility in Five Environmental Statutes" (Washington, D.C.: Environmental Law Institute, September 1995).

Environmental Law Institute, "Federal Oversight of Authorized State Environmental Programs: Reforming the System" (Washington, D.C.: Environmental Law Institute, September 1995).

Executive Session on Public Sector Performance Management, *Get Results Through Performance Management: An Open Memorandum to Government Executives* (Visions of Governance for the 21st Century Program: Kennedy School of Government, Harvard, 2001).

Frederickson, David, "The Potential of the Government Performance and Results Act as a Tool for Managing Third-Party Government" (Arlington, Va.: IBM Center for The Business of Government, 2001).

Ginsburg, Alan I., et al., "Lessons from the Wall Chart," *Educational Evaluation and Policy Analysis* (Spring 1988, Vol. 10, No.1), pp. 1-12.

Gormley, Jr., William T., and David L. Weimar, *Organizational Report Cards* (Cambridge: Harvard University Press, 1999), p. 39.

Grissmer, David, et al., *Improving Student Achievement; What State NAEP Test Scores Tell Us* (RAND Education, 2000).

Jerald, Craig D., *Dispelling the Myth Revisited: Preliminary Findings from a Nationwide Analysis of "High-Flying" Schools* (Washington, D.C.: The Education Trust, 2001).

Kenyon, Daphne, and John Kincaid, eds., *Competition Among States and Local Governments* (Washington, D.C.: The Urban Institute Press, 1991).

Kowaleski, Richard, "Using Outcome Information to Redirect Programs: A Case Study of the Coast Guard's Pilot Project Under the Government Performance and Results Act," United States Coast Guard Office of Marine Safety, Security and Environmental Protection, April 1996. (http://www.npr.gov/library/studies/uscgcase.pdf) [October 22, 2000].

Ladd, Helen F., *Holding Schools Accountable: Performance-Based Reform in Education* (Washington, D.C., Brookings, 1996).

Liner, Blaine, et al., *Making Results-Based State Government Work* (Washington, D.C.: The Urban Institute, 2001).

Markell, David, and Cliff Rechtschaffen, *Reinventing Environmental Enforcement and the State/Federal Relationship* (Washington, D.C.: Environmental Law Institute, 2003).

McLaughlin, Donald, "State Assessment and Data Policy Research," presentation for National Center for Education Statistics, Summer Conference 2002, Washington, D.C., July 25, 2002.

Metzenbaum, Shelley H., "Making Measurement Matter: The Challenge and Promise of Building a Performance-Focused Environmental Protection System" (Brookings Center for Public Management CPM 98-2, October 1998), p. 25.

Metzenbaum, Shelley H., "Measurement that Matters: Cleaning Up the Charles River," in Donald F. Kettl, ed., *Environmental Governance: A Report on the Next Generation of Environmental Policy* (Brookings Institution Press, 2001).

Metzenbaum, Shelley, "More Nutritious Beans," *Environmental Forum* (March/April 2002, Vol. 20, No. 2), pp. 19-41.

National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform* (Washington, D.C.: GPO, 1983) available at http://www.ed.gov/pubs/NatAtRisk/.

Radin, Beryl, "The Government Performance and Results Act (GPRA): Hydra-Headed Monster or Flexible Management Tool?," *Public Administration Review* (July/August 1998, Vol. 59, No. 4), pp. 307-315.

Ritter, Joyce N., "The History of Highway Statistics," *1994 Highway Statistics: 50th Anniversary Edition* (Washington, D.C.: U. S. Department of Transportation, 1994).

Seely, Bruce E., *Building the American Highway System: Engineers as Policy Makers* (Philadelphia: Temple University Press, 1987).

Tyack, David, and Larry Cuban, *Tinkering Toward Utopia: A Century of Public School Reform* (Cambridge: Harvard University Press, 1995).

U.S. Department of Education, Office of the Under Secretary and Office of Elementary and Secondary Education, "School Improvement Report: Executive Order on Actions for Turning Around Low-Performing Schools" (Washington, D.C.: January 2001).

U.S. Environmental Protection Agency, "Report of the Task Force to Enhance State Capacity: Strengthening Environmental Management in the States" (Office of the Administrator, EPA 270-R-93-001, July 1993).

U.S. General Accounting Office, "Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information," (Washington, D.C.: GAO/GGD-99-139, July 1999).

U.S. General Accounting Office, "Title I: Education Needs to Monitor States' Scoring of Assessments" (Washington, D.C.: GAO-02-393, April 1, 2002).

U.S. General Accounting Office, "Transportation Infrastructure: States' Implementation of Transportation Management Systems" (Washington, D.C.: GAO/RCED-97-32, January 1997).

Sources

The author wishes to express her great appreciation to the many people who generously shared their time and expertise, some of whom are listed here and others who preferred not to be identified by name. I am truly indebted to them all for their assistance.

EPA

Steve Brown, Executive Director, Environmental Council of the States; **Michael Cook,** Director, Office of Wastewater Management; **Michael Shapiro,** Deputy Regional Administrator, Office of Solid Waste and Emergency Response; **John Seitz,** Office of Air Quality Planning and Standards; **Robert Springer,** Region 5; **Bill Becker,** State and Territorial Air Pollution Program Administrators/ Association of Local Air Pollution Control Officials; and selected state officials who are members of the Environmental Compliance Consortium and in other state agencies. Much of the description of the EPA experience is based on discussions with hundreds of EPA and state staff over the past 10 years, including during the author's tenure as Associate Administrator of EPA for Regional Operations and State/Local Relations from 1993 to 1997, Undersecretary of the Massachusetts Executive Office of Environmental Affairs prior to that, and current work as the director of an organization working with state environmental agencies.

Transportation

Mort Downey, Principal Consultant, PBConsult, former Deputy Secretary of the U.S. Department of Transportation; **Tony Kane,** Director, Engineering and Technical Services, American Association of State Highway and Transportation, former Administrator of the Federal Highway Administration; **Richard Weingroff,** Information Liaison Specialist, FHWA. Over a dozen other officials in the central and regional offices of the U.S. Department of Transportation and state transportation agencies were also incredibly helpful.

Education

Allen Ginsberg, Director, Office of Planning and Educational Statistics; **Jack Jennings,** former congressional staff; **David Bushnell,** Bowie State University; and several others currently or formerly in the U.S. Department of Education and nonprofit organizations working in the education field.

General

John Kincaid, Lafayette University; **Barry White,** Council for Excellence in Government; **Bruce McDowell,** National Academy of Public Administration; **Paul Posner,** U.S. General Accounting Office; and **Jonathan Breul,** IBM Center for The Business of Government.

A special note of thanks to **Mark Abramson** and **John Kamensky** of the IBM Center for The Business of Government for sharing their substantive knowledge and continuing advice on the report itself.

ABOUT THE AUTHOR

Shelley H. Metzenbaum is a Visiting Professor at the University of Maryland School of Public Affairs and Executive Director of the Environmental Compliance Consortium, a collaborative effort of state environmental protection agencies to improve the measurement and management of their programs.

Metzenbaum is the faculty chair of the University of Maryland School of Public Affairs' executive education program for the U.S. Food and Drug Administration. Prior to that, she served as adjunct faculty at the Brookings Institution, running the "Science and Technology Policy" and "Managing for Results" programs. From 1998 to 2000, she headed the executive session on public sector performance management at Harvard's Kennedy School of Government. She currently serves as senior consultant to the Facility Reporting Project, a multi-stakeholder project to develop generally accepted, credible, and comparable sustainability metrics for regulated facilities.

From 1993 to 1997, Metzenbaum served as associate administrator of the U.S. Environmental Protection Agency for Regional Operations and State/Local Relations. As associate administrator, she managed the design and implementation of the National Environmental Performance Partnership System (NEPPS) and the Sustainable Development Challenge Grant program. Previously, Metzenbaum served as undersecretary of the Massachusetts Executive Office of Environmental Affairs (EOEA.) Before serving as undersecretary, Metzenbaum was director of capital budgeting for the Commonwealth of Massachusetts.

Metzenbaum's teaching, writing, and consulting emphasize ways governments can make performance measurement more useful, with a special emphasis on ways information can improve the effectiveness and efficiency of the environmental protection system. Metzenbaum holds a Ph.D. in public policy from the Kennedy School of Government, Harvard University. She is a fellow of the National Academy of Public Administration and a principal of the Council for Excellence in Government.

KEY CONTACT INFORMATION

To contact the author:

Shelley H. Metzenbaum
Visiting Professor
Maryland School of Public Affairs
University of Maryland
2101 Van Munching Hall
College Park, MD 20742
(978) 371-3099

e-mail: smetzenbaum@aics.net

E-GOVERNMENT

Supercharging the Employment Agency: An Investigation of the Use of Information and Communication Technology to Improve the Service of State Employment Agencies (December 2000)

Anthony M. Townsend

Assessing a State's Readiness for Global Electronic Commerce: Lessons from the Ohio Experience (January 2001)

J. Pari Sabety
Steven I. Gordon

Privacy Strategies for Electronic Government (January 2001)

Janine S. Hiller
France Bélanger

Commerce Comes to Government on the Desktop: E-Commerce Applications in the Public Sector (February 2001)

Genie N. L. Stowers

The Use of the Internet in Government Service Delivery (February 2001)

Steven Cohen
William Eimicke

State Web Portals: Delivering and Financing E-Service (January 2002)

Diana Burley Gant
Jon P. Gant
Craig L. Johnson

Internet Voting: Bringing Elections to the Desktop (February 2002)

Robert S. Done

Leveraging Technology in the Service of Diplomacy: Innovation in the Department of State (March 2002)

Barry Fulton

Federal Intranet Work Sites: An Interim Assessment (June 2002)

Julianne G. Mahler
Priscilla M. Regan

The State of Federal Websites: The Pursuit of Excellence (August 2002)

Genie N. L. Stowers

State Government E-Procurement in the Information Age: Issues, Practices, and Trends (September 2002)

M. Jae Moon

Preparing for Wireless and Mobile Technologies in Government (October 2002)

Ai-Mei Chang
P. K. Kannan

Public-Sector Information Security: A Call to Action for Public-Sector CIOs (October 2002, 2nd ed.)

Don Heiman

The Auction Model: How the Public Sector Can Leverage the Power of E-Commerce Through Dynamic Pricing (November 2002, 2nd ed.)

David C. Wyld

The Promise of E-Learning in Africa: The Potential for Public-Private Partnerships (January 2003)

Norman LaRocque
Michael Latham

Digitally Integrating the Government Supply Chain: E-Procurement, E-Finance, and E-Logistics (February 2003)

Jacques S. Gansler
William Lucyshyn
Kimberly M. Ross

Using Technology to Increase Citizen Participation in Government: The Use of Models and Simulation (April 2003)

John O'Looney

Services Acquisition for America's Navy: Charting a New Course for SeaPort (June 2003)

David C. Wyld

FINANCIAL MANAGEMENT

Credit Scoring and Loan Scoring: Tools for Improved Management of Federal Credit Programs (July 1999)

Thomas H. Stanton

Using Activity-Based Costing to Manage More Effectively (January 2000)

Michael H. Granof
David E. Platt
Igor Vaysman

Audited Financial Statements: Getting and Sustaining "Clean" Opinions (July 2001)

Douglas A. Brook

An Introduction to Financial Risk Management in Government (August 2001)

Richard J. Buttimer, Jr.

Understanding Federal Asset Management: An Agenda for Reform (July 2003)

Thomas H. Stanton

Efficiency Counts: Developing the Capacity to Manage Costs at Air Force Materiel Command (August 2003)

Michael Barzelay
Fred Thompson

MARKET-BASED GOVERNMENT

Determining a Level Playing Field for Public-Private Competition (November 1999)

Lawrence L. Martin

Implementing State Contracts for Social Services: An Assessment of the Kansas Experience (May 2000)

Jocelyn M. Johnston
Barbara S. Romzek

A Vision of the Government as a World-Class Buyer: Major Procurement Issues for the Coming Decade (January 2002)

Jacques S. Gansler

Contracting for the 21st Century: A Partnership Model (January 2002)

Wendell C. Lawther

Franchise Funds in the Federal Government: Ending the Monopoly in Service Provision (February 2002)

John J. Callahan

Making Performance-Based Contracting Perform: What the Federal Government Can Learn from State and Local Governments (November 2002, 2nd ed.)

Lawrence L. Martin

Moving to Public-Private Partnerships: Learning from Experience around the World (February 2003)

Trefor P. Williams

IT Outsourcing: A Primer for Public Managers (February 2003)

Yu-Che Chen
James Perry

The Procurement Partnership Model: Moving to a Team-Based Approach (February 2003)

Kathryn G. Denhardt

Moving Toward Market-Based Government: The Changing Role of Government as the Provider (June 2003)

Jacques S. Gansler

HUMAN CAPITAL MANAGEMENT

Profiles in Excellence: Conversations with the Best of America's Career Executive Service (November 1999)

Mark W. Huddleston

Reflections on Mobility: Case Studies of Six Federal Executives (May 2000)

Michael D. Serlin

Managing Telecommuting in the Federal Government: An Interim Report (June 2000)

Gina Vega
Louis Brennan

Using Virtual Teams to Manage Complex Projects: A Case Study of the Radioactive Waste Management Project (August 2000)

Samuel M. DeMarie

A Learning-Based Approach to Leading Change (December 2000)

Barry Sugarman

Labor-Management Partnerships: A New Approach to Collaborative Management (July 2001)

Barry Rubin
Richard Rubin

Winning the Best and Brightest: Increasing the Attraction of Public Service (July 2001)

Carol Chetkovich

A Weapon in the War for Talent: Using Special Authorities to Recruit Crucial Personnel (December 2001)

Hal G. Rainey

A Changing Workforce: Understanding Diversity Programs in the Federal Government (December 2001)

Katherine C. Naff
J. Edward Kellough

Life after Civil Service Reform: The Texas, Georgia, and Florida Experiences (October 2002)

Jonathan Walters

The Defense Leadership and Management Program: Taking Career Development Seriously (December 2002)

Joseph A. Ferrara
Mark C. Rom

The Influence of Organizational Commitment on Officer Retention: A 12-Year Study of U.S. Army Officers (December 2002)

Stephanie C. Payne
Ann H. Huffman
Trueman R. Tremble, Jr.

Human Capital Reform: 21st Century Requirements for the United States Agency for International Development (March 2003)

Anthony C. E. Quainton
Amanda M. Fulmer

Modernizing Human Resource Management in the Federal Government: The IRS Model (April 2003)

James R. Thompson
Hal G. Rainey

Mediation at Work: Transforming Workplace Conflict at the United States Postal Service (October 2003)

Lisa B. Bingham

Growing Leaders for Public Service (November 2003)

Ray Blunt

MANAGING FOR PERFORMANCE AND RESULTS

Corporate Strategic Planning in Government: Lessons from the United States Air Force (November 2000)

Colin Campbell

Using Evaluation to Support Performance Management: A Guide for Federal Executives (January 2001)

Kathryn Newcomer
Mary Ann Scheirer

Managing for Outcomes: Milestone Contracting in Oklahoma (January 2001)

Peter Frumkin

The Challenge of Developing Cross-Agency Measures: A Case Study of the Office of National Drug Control Policy (August 2001)

Patrick J. Murphy
John Carnevale

The Potential of the Government Performance and Results Act as a Tool to Manage Third-Party Government (August 2001)

David G. Frederickson

Using Performance Data for Accountability: The New York City Police Department's CompStat Model of Police Management (August 2001)

Paul E. O'Connell

Moving Toward More Capable Government: A Guide to Organizational Design (June 2002)

Thomas H. Stanton

Performance Management: A "Start Where You Are, Use What You Have" Guide (October 2002)

Chris Wye

The Baltimore CitiStat Program: Performance and Accountability(May 2003)

Lenneal J. Henderson

How Federal Programs Use Outcome Information: Opportunities for Federal Managers (May 2003)

Harry P. Hatry
Elaine Morley
Shelli B. Rossman
Joseph S. Wholey

Linking Performance and Budgeting: Opportunities in the Federal Budget Process (October 2003)

Philip G. Joyce

Strategies for Using State Information: Measuring and Improving Program Performance (December 2003)

Shelley H. Metzenbaum

INNOVATION

Managing Workfare: The Case of the Work Experience Program in the New York City Parks Department (June 1999)

Steven Cohen

New Tools for Improving Government Regulation: An Assessment of Emissions Trading and Other Market-Based Regulatory Tools (October 1999)

Gary C. Bryner

Religious Organizations, Anti-Poverty Relief, and Charitable Choice: A Feasibility Study of Faith-Based Welfare Reform in Mississippi (November 1999)

John P. Bartkowski
Helen A. Regis

Business Improvement Districts and Innovative Service Delivery (November 1999)

Jerry Mitchell

An Assessment of Brownfield Redevelopment Policies: The Michigan Experience (November 1999)

Richard C. Hula

San Diego County's Innovation Program: Using Competition and a Whole Lot More to Improve Public Services (January 2000)

William B. Eimicke

Innovation in the Administration of Public Airports (March 2000)

Scott E. Tarry

Entrepreneurial Government: Bureaucrats as Businesspeople (May 2000)

Anne Laurent

Rethinking U.S. Environmental Protection Policy: Management Challenges for a New Administration (November 2000)

Dennis A. Rondinelli

The Challenge of Innovating in Government (February 2001)

Sandford Borins

Understanding Innovation: What Inspires It? What Makes It Successful? (December 2001)

Jonathan Walters

Government Management of Information Mega-Technology: Lessons from the Internal Revenue Service's Tax Systems Modernization (March 2002)

Barry Bozeman

Advancing High End Computing: Linking to National Goals (September 2003)

Juan D. Rogers
Barry Bozeman

NETWORKS, COLLABORATION, AND PARTNERSHIPS

Leveraging Networks to Meet National Goals: FEMA and the Safe Construction Networks (March 2002)

William L. Waugh, Jr.

21st-Century Government and the Challenge of Homeland Defense (June 2002)

Elaine C. Kamarck

Assessing Partnerships: New Forms of Collaboration (March 2003)

Robert Klitgaard
Gregory F. Treverton

Leveraging Networks: A Guide for Public Managers Working across Organizations (March 2003)

Robert Agranoff

Extraordinary Results on National Goals: Networks and Partnerships in the Bureau of Primary Health Care's 100%/0 Campaign (March 2003)

John Scanlon

Public-Private Strategic Partnerships: The U.S. Postal Service-Federal Express Alliance (May 2003)

Oded Shenkar

The Challenge of Coordinating "Big Science" (July 2003)

W. Henry Lambright

Communities of Practice: A New Tool for Government Managers (November 2003)

William M. Snyder
Xavier de Souza Briggs

TRANSFORMING ORGANIZATIONS

The Importance of Leadership: The Role of School Principals (September 1999)

Paul Teske
Mark Schneider

Leadership for Change: Case Studies in American Local Government (September 1999)

Robert B. Denhardt
Janet Vinzant Denhardt

Managing Decentralized Departments: The Case of the U.S. Department of Health and Human Services (October 1999)

Beryl A. Radin

Transforming Government: The Renewal and Revitalization of the Federal Emergency Management Agency (April 2000)

R. Steven Daniels
Carolyn L. Clark-Daniels

Transforming Government: Creating the New Defense Procurement System (April 2000)

Kimberly A. Harokopus

Trans-Atlantic Experiences in Health Reform: The United Kingdom's National Health Service and the United States Veterans Health Administration (May 2000)

Marilyn A. DeLuca

Transforming Government: The Revitalization of the Veterans Health Administration (June 2000)

Gary J. Young

The Challenge of Managing Across Boundaries: The Case of the Office of the Secretary in the U.S. Department of Health and Human Services (November 2000)

Beryl A. Radin

Creating a Culture of Innovation: 10 Lessons from America's Best Run City (January 2001)

Janet Vinzant Denhardt
Robert B. Denhardt

Transforming Government: Dan Goldin and the Remaking of NASA (March 2001)

W. Henry Lambright

Managing Across Boundaries: A Case Study of Dr. Helene Gayle and the AIDS Epidemic (January 2002)

Norma M. Riccucci

Managing "Big Science": A Case Study of the Human Genome Project (March 2002)

W. Henry Lambright

The Power of Frontline Workers in Transforming Government: The Upstate New York Veterans Healthcare Network (April 2003)

Timothy J. Hoff

Making Public Sector Mergers Work: Lessons Learned (August 2003)

Peter Frumkin

Efficiency Counts: Developing the Capacity to Manage Costs at Air Force Materiel Command (August 2003)

Michael Barzelay
Fred Thompson

Managing the New Multipurpose, Multidiscipline University Research Centers: Institutional Innovation in the Academic Community (November 2003)

Barry Bozeman
P. Craig Boardman

HEALTHCARE

The Power of Frontline Workers in Transforming Healthcare Organizations: The Upstate New York Veterans Healthcare Network (December 2003)

Timothy J. Hoff

IT Outsourcing: A Primer for Healthcare Managers (December 2003)

Yu-Che Chen
James Perry

BOOKS*

E-Government 2001
(Rowman & Littlefield Publishers, Inc., 2001)
Mark A. Abramson and Grady E. Means, editors

E-Government 2003
(Rowman & Littlefield Publishers, Inc., 2002)
Mark A. Abramson and Therese L. Morin, editors

Human Capital 2002
(Rowman & Littlefield Publishers, Inc., 2002)
Mark A. Abramson and Nicole Willenz Gardner, editors

Innovation
(Rowman & Littlefield Publishers, Inc., 2002)
Mark A. Abramson and Ian Littman, editors

Leaders
(Rowman & Littlefield Publishers, Inc., 2002)
Mark A. Abramson and Kevin M. Bacon, editors

Managing for Results 2002
(Rowman & Littlefield Publishers, Inc., 2001)
Mark A. Abramson and John Kamensky, editors

Memos to the President: Management Advice from the Nation's Top Public Administrators
(Rowman & Littlefield Publishers, Inc., 2001)
Mark A. Abramson, editor

New Ways of Doing Business
(Rowman & Littlefield Publishers, Inc., 2003)
Mark A. Abramson and Ann M. Kieffaber, editors

The Procurement Revolution
(Rowman & Littlefield Publishers, Inc., 2003)
Mark A. Abramson and Roland S. Harris III, editors

Transforming Government Supply Chain Management
(Rowman & Littlefield Publishers, Inc., 2003)
Jacques S. Gansler and Robert E. Luby, Jr., editors

Transforming Organizations
(Rowman & Littlefield Publishers, Inc., 2001)
Mark A. Abramson and Paul R. Lawrence, editors

* Available at bookstores, online booksellers, and from the publisher (www.rowmanlittlefield.com or 800-462-6420).

About IBM Business Consulting Services

With consultants and professional staff in more than 160 countries globally, IBM Business Consulting Services is the world's largest consulting services organization. IBM Business Consulting Services provides clients with business process and industry expertise, a deep understanding of technology solutions that address specific industry issues, and the ability to design, build and run those solutions in a way that delivers bottom-line business value. For more information visit www.ibm.com/bcs.

About the IBM Center for The Business of Government

Through research stipends and events, the IBM Center for The Business of Government stimulates research and facilitates discussion on new approaches to improving the effectiveness of government at the federal, state, local, and international levels.

The Center is one of the ways that IBM seeks to advance knowledge on how to improve public sector effectiveness. The IBM Center focuses on the future of the operation and management of the public sector.

For additional information, contact:
Mark A. Abramson
Executive Director
IBM Center for The Business of Government
1301 K Street, NW
4th Floor, West Tower
Washington, DC 20005
(202) 515-4504, fax: (202) 515-4375

e-mail: businessofgovernment@us.ibm.com
website: www.businessofgovernment.org