

Public

Policymaking

An Introduction

SIXTH EDITION

JAMES E. ANDERSON

Texas A&M University

HOUGHTON MIFFLIN COMPANY

BOSTON NEW YORK

- "Duty, Fear, and Tax Compliance: The Heuristic Basis of Citizen Behavior," *American Journal of Political Science*, Vol. 39 (May 1995), pp. 490-512.
82. This discussion depends heavily on ideas in my "Public Economic Policy and the Problems of Compliance: Notes for Research," *Houston Law Review*, IV (Spring-Summer 1966), pp. 62-72.
 83. Tom R. Tyler, *Why People Obey the Law* (New Haven: Yale University Press, 1996), p. 178.
 84. Anne Schneider and Helen Ingram, "Behavioral Theories in Policy Designs." Unpublished paper presented at the Midwest Political Association meeting (1989).
 85. Herbert A. Simon, Donald Smithburg, and Victor Thompson, *Public Administration* (New York: Knopf, 1950), p. 479.
 86. *Los Angeles Times*, March 1, 2002, part 3, p. 1.
 87. David Cay Johnston, "Fearing for Jobs, I.R.S. Workers Relax Efforts to Get Unpaid Taxes," *New York Times*, May 18, 1999, pp. A1, C11.
 88. *Newsday*, January 22, 2002, p. A26.
 89. John D. McKinnon, "IRS Rides the Ups and Downs of Congressional Whims," *Wall Street Journal*, April 8, 2002, p. A28.
 90. Robert Lane, *The Regulation of Businessmen* (New Haven: Yale University Press, 1954), pp. 69-70.
 91. 370 U.S. 421 (1962).
 92. See John A. Murley, "School Prayer: Free Exercise of Religion or Establishment of Religion," in Raymond Tatalovich and Byron W. Daynes, eds., *Social Regulatory Policy* (Boulder, CO: Westview Press, 1989), chap. 1.
 93. *Santa Fe Independent School District v. Doe*, 530 U.S. 1 (2000).
 94. Timothy B. Clark, "Honesty May Become the Best Tax Policy If Tax Compliance Bill Becomes Law," *National Journal*, Vol. 14 (July 24, 1982), pp. 1292-1296.
 95. Marshall B. Clinard, *Sociology of Deviant Behavior* (New York: Holt, Rinehart and Winston, 1957), pp. 168-171.
 96. *Ibid.*
 97. Robert E. Lane, "Why Business Men Violate the Law," *Journal of Criminal Law, Criminology, and Police Science*, XLIV (1953), pp. 151, 154-160.
 98. John G. Fuller, *The Gentlemen Conspirators: The Story of the Price-Fixers in the Electrical Industry* (New York: Grove Press, 1962), pp. 88, 109-110.
 99. G. Pascal Zachary, "The Minimum Wage Law Is Frequently Ignored in Some Industries," *Wall Street Journal*, May 20, 1996, p. 1. See also *idem.*, "Many Firms Refuse to Pay Overtime, Employees Complain," *Wall Street Journal*, June 24, 1996, p. 1.
 100. Lane, *op. cit.*, chap. 5.
 101. Kerwin, *op. cit.*, p. 81.
 102. Simon, Smithburg, and Thompson, *op. cit.*, p. 457.
 103. Roger G. Noll, *Reforming Regulation* (Washington, DC: Brookings Institution, 1971), p. 65.
 104. *Wall Street Journal*, May 19, 1977, p. 40; and Timothy B. Clark, "What's All the Uproar Over OSHA's Nit-Picking Rules?" *National Journal*, Vol. 10 (October 7, 1978), pp. 1594-1596.
 105. General Accounting Office, *Environmental Enforcement: Penalties May Not Recover Economic Benefits Gained by Violators* (Washington, DC: USGAO, June 1991), pp. 5-10.

Policy Impact, Evaluation, and Change



When it is viewed as a cycle or sequential pattern of functional activities, the final phase of the policy process is evaluation. As much an art as a science, policy evaluation involves the estimation, appraisal, or assessment of a policy, its content, implementation, goal attainment, and other effects. The policy evaluator wants to know whether and to what extent a policy has accomplished its goals or whether it has had other effects, intended or unintended.

Policy evaluation may also seek to identify factors that contributed to the success or failure of a policy. This, in turn, may lead to recycling of the policy process (problem definition, formulation, adoption, and so on) in order to continue, modify, strengthen, or terminate the policy. Put differently, information gained through evaluation feeds back into the policy process.

As a functional activity, evaluation can occur at any point in the policy process, not simply after some effort has been made to implement a policy. Thus, attempts are made to determine prospectively—to estimate or predict—the likely effects, or the costs and benefits, of policy alternatives prior to their adoption. Typically, however, policy evaluation "looks backward" to what has happened whereas the other stages of the policy process look forward to attaining a goal.¹

Before proceeding further in examining evaluation, we need to pause and look at policy impacts, at the general sorts of effects or consequences that may ensue from policies.



Policy Impact

To begin, there is a need to draw a distinct line between *policy outputs* and *policy outcomes*. *Policy outputs* are the things actually done by agencies in pursuance of policy decisions and statements. The concept of outputs focuses one's attention on such matters as amounts of taxes collected, miles of highways built, welfare benefits paid, price-fixing agreements prosecuted, traffic fines collected, or foreign-aid projects undertaken. Outputs usually can be

readily counted, totaled, and statistically analyzed. Examining outputs may indicate, or seem to indicate, that a lot is being done to implement a policy. Such activity, however, sometimes amounts to nothing more than what Professor William T. Gormley Jr. calls "bean counting." Agencies, under pressure from legislators, interest groups, and others to demonstrate results, "may focus on outputs, not outcomes, in order to generate statistics that create the illusion of progress."² If the percentage of students graduating from universities in a state increases, does this tell us anything about the quality of education that they are receiving?

Policy outcomes (sometimes called "results"), in contrast, are the consequences for society, intended and unintended, that stem from deliberate governmental action or inaction. Social-welfare policies can be used to illustrate this concept. It is fairly easy to measure welfare-policy outputs such as amounts of benefits paid, average level of benefits, and number of people assisted. But what are the outcomes, or societal consequences, of these actions? Do they increase personal security and contentment? Do they reduce individual initiative? Did Aid to Families with Dependent Children (AFDC) have the effect of encouraging promiscuity and illegitimacy, or teenage pregnancies, as some alleged? Do welfare programs help keep the poor quiescent, as others contend?³ Questions such as these, which are tough to answer, direct our attention to the societal effects of policies. Among other things, policy students should want to know whether policies are accomplishing their intended purposes, whether society is changing as a consequence of policy actions and not because of other factors such as private economic decisions, and whether it is changing as intended or in other ways. Policy impacts are an amalgam of outputs and outcomes.

The impact of a policy may have several dimensions, all of which should be taken into account either in the conduct of a formal evaluation or in the course of an informal appraisal of the policy.⁴ They include the following:

1. Policies affect the public problem at which they are directed and the people involved. The target populations whom the policy is intended to affect must be defined, whether they are the poor, small-business people, disadvantaged schoolchildren, petroleum producers, or whoever. The intended effect of the policy must then be determined. If it is an antipoverty program, is its purpose to raise the income of the poor, to increase their opportunities for employment, or to change their attitudes and behavior? If some combination of such purposes is intended, analysis becomes more complicated because priorities should be assigned to the various intended effects. Typically, policies accomplish at least a portion of their goals or objectives.

Further, a policy may have either intended or unintended (unforeseen or unplanned) consequences, or even both. A welfare program, for example, may improve the income situation of the benefited groups, as intended by its proponents, but what impact does it have on their initiative to seek employment? Does it decrease this initiative, as some have contended? A public-housing program may improve the housing situation of urban blacks, but it may also

be so administered as to contribute to racial segregation in housing. Agricultural income-support programs, intended to improve farmers' incomes, also may lead to overproduction of the supported commodities, to higher food prices for consumers, or to an unearned increase in land values.

A good illustration of a policy with unintended consequences is the 1970 legislation prohibiting the broadcasting of cigarette advertising on radio and television. This ad ban was widely viewed as a legislative victory for the anti-smoking forces. However, the ban also eliminated the need for broadcasters, under the Federal Communications Commission's fairness doctrine (which was repealed in the mid-1980s), to donate airtime to antismoking groups on the controversial issue of smoking. Research indicates that the antismoking messages prepared by these groups, which contained grim or unappealing portrayals of smoking, had a substantial deterrent effect upon cigarette consumption. The antismoking ad campaign, however, depended heavily upon donated time. As a consequence, after the ban on cigarette ads went into effect, most of the antismoking ads were also eliminated. The short-term effect was clearly a significant increase in smoking, obviously not what the proponents of the ban had intended. Although the long-run effects are less clear, "the weight of the evidence seems to favor the conclusion that the ad ban was myopic policy."⁵

2. Policies may affect situations or groups other than those at which they are directed. These are variously called third-party effects, spillover effects, or externalities.⁶ The construction of urban interstate highways are of much benefit to motorists and trucking companies. However, they also cause inconvenience, disorder, and social disruption for the neighborhoods through which they run. Clear-cutting in national forests, which is of benefit to timber companies and in line with the perspective of those who view forests as tree farms, is profoundly disturbing to environmentalists, nature lovers, and many sportsmen because it results in the destruction of wildlife habitat, the loss of aesthetic value, and the siltation of trout streams.

These two examples portray negative externalities, but externalities may also be positive. Public-education programs not only educate students; they also provide employers with a more capable workforce and the community with better-informed citizens. A strong line of research supports a correlation between education and support for democracy. Those who contend that only those who have children in public schools should contribute toward their support ignore such positive externalities. Although pollution-control programs impose costs on many industries, they are a boon to the manufacturers of pollution-control equipment. Many of the outcomes of public policies can be most meaningfully understood as externalities that impose costs or provide benefits for third parties.

3. Policies have consequences for future as well as current conditions; for some policies most of their benefits or some of their costs may occur in the far future. Was the Head Start program—a preschool education program for the poor—supposed to improve participating children's short-term cognitive abilities or their long-range development and earning capacity? Did regulation of

the field price of natural gas, a policy initiated in the 1950s and extending into the 1980s, really produce a shortage of gas in the 1970s, as some contended (notably petroleum-industry officials and their supporters, who had long been critics of price regulation)?

The future effects of some policies may be very diffuse or uncertain. Assuming that patent and copyright policies do indeed stimulate invention and creativity, and that these activities in turn enhance economic growth and societal development, how does one measure their benefits, either qualitatively or quantitatively? Again, how does one appraise (with reasonable objectivity) the effects of the National Foundation on the Arts and the Humanities' support for literacy, artistic, and museum activities? Would the elimination of such policies as these have deleterious consequences for American society?

4. Just as policies have positive effects or benefits, they also entail costs. Economists seem never to tire of telling us that there is no such thing as a free lunch. Costs come in different forms. First, there are the direct costs for the governmental implementation of a policy or program. These are usually fairly easy to calculate, whether stated as the actual amount of money spent on the program, its share of total governmental expenditures, or the proportion of the gross domestic product devoted to it. Budgeting documents will yield such figures. If, however, a governmental expenditure serves multiple purposes, such as operating the space program and developing new technology, the allocation of costs becomes more perplexing.

Direct costs also include private expenditures that are necessary in order to comply with public policies, such as those on industrial health and safety and environmental-pollution control. These may be more difficult to discover or calculate. Moreover, it is possible that some companies would have installed protective devices in the absence of policy, whether out of a desire to serve the public interest or to protect themselves from lawsuits. Should their costs then be assigned to the policy? In the absence of governmental subsidies, the direct costs of complying with regulatory policies initially fall primarily on the regulated, who have an ideological incentive to inflate claimed costs, deliberately making the policies appear more burdensome. Ultimately, of course, many compliance costs are likely to be shifted to consumers in the form of higher prices for goods and services.

This brings us to the matter of indirect costs. Public policies may cause reduced production, higher prices, or mental anguish or distress. Expenditures to meet coal-mine-safety requirements may cause a reduction in mine output. People called for jury services typically receive compensation that does no more than cover commuting costs. The consequence is lost wages, or lost production, or both. Many public policies have indirect or social costs of these sorts. Mental anguish may occur when one's hometown is flooded by a new impoundment. Financial compensation may be paid for one's childhood home, but what of the loss that occurs because, being under forty feet of water, it can no longer be visited? The concept of indirect cost often calls for putting a price tag on intangibles. They tend to defy measurement.

Finally, there are opportunity costs, a concept that rests upon the facts that we cannot do everything at once and that resources used for one purpose (e.g., flood control) cannot be used for another purpose (e.g., public housing). "Opportunity cost is a decision making rather than an accounting concept."⁷ It focuses attention on what one has to give up (or, alternatively, what one will gain) if resources are used for one purpose rather than another. Thus, economists argue that the United States's all-volunteer army will not attract needed recruits if the opportunity costs of military service are kept below those of civilian life. When the ranks of the military were largely filled by draftees, less concern had to be given to opportunity costs and the compensation of service personnel.⁸

5. The effects of policies and programs may be either material (tangible) or symbolic (intangible). The consequences of symbols are both important and hard to measure. Symbolic outputs, according to political scientists Gabriel A. Almond and G. Bingham Powell, "include affirmations of values by elites; displays of flags, troops, and military ceremony; visits by royalty or high officials; and statements of policy or intent by political leaders." Consequently, they "are highly dependent on tapping popular beliefs, attitudes, and aspirations for their effectiveness."⁹

Symbolic policy outputs produce no readily discernible changes in societal conditions. No one eats better because of a Memorial Day parade or a stirring speech by a high public official on the virtues of free enterprise, however ideologically or emotionally satisfying such actions may be for many people. More to the point, however, policy actions ostensibly directed toward meeting material wants or needs may turn out in practice to be more symbolic than material in their effect.

This shift in policy tone is well illustrated by the Fair Housing Act of 1968. Enacted by Congress in part because of pressure created by the assassination of Dr. Martin Luther King Jr., the 1968 law prohibited discrimination in the rental or sale of housing because of race, color, religion, sex, or national origin. However, the Department of Housing and Urban Development (HUD), which was assigned primary responsibility for its enforcement, was authorized only to mediate disputes between a person who thought he or she had been discriminated against and the renter or seller. The Justice Department, in turn, could not act unless it found a "pattern or practice" of discrimination. As a consequence of these weak enforcement provisions, the Fair Housing Act in practice did not live up to its promise. The act, which one member of Congress called a "toothless tiger," was of little use in preventing discrimination in housing.

The Congress in 1988 reached a compromise agreement on legislation to strengthen enforcement of the Fair Housing Act. Now a person who believes he or she has been discriminated against can file a complaint with HUD, which, if it cannot settle the dispute, can issue a charge of housing discrimination. At this point either party to the dispute can choose to have it decided by either a federal district court or an administrative hearing. If either party

chooses to go to the federal court, that choice prevails. This procedure was intended to put some real "teeth" into the enforcement of the act and give it material rather than merely symbolic effect.

Other public policies that appear to promise more symbolically than their implementation yields in material benefits include antitrust activity, public-utility rate regulation, and equal employment opportunity. Even though the actual effect of a policy may be considerably less than is intended or desired, it nonetheless may have significant consequences for society. An antipoverty program that falls short of its mark may provide some assurance to people that the government cares about poor people and wants to reduce poverty. Legislation on equal employment opportunity informs people that their government, officially at least, does not condone discrimination in hiring on the basis of race, sex, or nationality. Apart from the effects such policies have on societal conditions, they may contribute to the maintenance of social order, support for government, and personal self-esteem, which are not inconsequential considerations.

The analysis and evaluation of public policy is usually focused upon what governments actually do, why, and with what material effects. We should not, however, neglect the symbolic aspects of government, despite their intangible and nebulous nature. The rhetoric of government—what governments say, or appear to say—is clearly a necessary and proper subject for the policy analyst.



Policy Evaluation

Policy evaluation, as a functional activity, is as old as policy itself. Legislators, administrators, judges, pressure-group officials, media commentators, and citizens have always made judgments about the worth or effects of particular policies, programs, and projects. The Boston Tea Party, for instance, was the result of an unfavorable evaluation of one of King George's tax policies for the colonies. Many of these judgments have been of an impressionistic or intuitive variety, based at best on anecdotal or fragmentary evidence, and strongly influenced by ideological, partisan, or idiosyncratic valuational criteria. Thus, staunch conservatives may regard public-housing programs as socialistic and in need of repeal, regardless of their causes, necessity, or consequences; or Democrats may support higher taxes on corporations and the rich because they believe this will enhance their electoral opportunities. Unemployment compensation may be deemed a "bad" program because the evaluator claims to know "a lot of people" who improperly receive benefits. Apocryphal stories about "welfare queens" who drive Cadillacs to collect their welfare checks are commonplace among welfare critics. Most of us are familiar with this style of policy evaluation and have perhaps enjoyed doing a bit of it ourselves. Much conflict results from this sort of evaluation, however, because different evaluators, depending upon the values or criteria they employ, may reach sharply divergent conclusions on the merits of the same policy.

Another form of policy evaluation centers on process, on the operation or administration of a policy or program.¹⁰ Although the terms *policy* and *program*

are often used interchangeably, a line can be drawn between them. *Policy* is defined in this book as a course of action followed by government in dealing with some problem or matter of concern. A policy may spawn a variety of programs, a *program* being defined as a set of rules, routines, and resources intended to carry out a policy or a portion thereof. Programs, which give "direct, specific, and concrete expression" to the policies on which they are based, sometimes are marked out in legislation. More likely, however, they will be the handiwork of administrative officials. Thus, the Antitrust Division has created several programs to enforce the antitrust laws.

A program, or several programs, conceivably could be eliminated without impairment of the underlying policy. In 2003, the Bush administration proposed eliminating several fairly specifically focused education programs because they were deemed unnecessary, duplicatory, or replaceable. However, the broad, underlying policy of federal support for public elementary and secondary education remained firmly in place. Most of the targeted programs were retained by Congress.

Questions asked about how a policy or program is being run may include: What are its financial costs? Who receives benefits (payments or services) and in what amounts? Is there unnecessary overlap with or duplication of other programs? Are legislatively prescribed standards and procedures being followed? Is the program honestly administered? This kind of evaluation, which may involve much monitoring of agencies and their officials, will tell us something about whether there is honesty or efficiency in the conduct of a program, but, like the first kind of evaluation, it will probably yield little or nothing in the way of hard information concerning the societal effects (outcomes) of a program. On the other hand, process evaluation is often helpful to program managers wanting to improve the administration of their programs and reduce their potential for political criticism.

A third, comparatively new type of policy evaluation, which has been getting increasing attention in the national government in recent decades, is the systematic and intendedly objective evaluation of programs. This form of evaluation, which I will refer to as *systematic evaluation*, employs social science methodology to measure the societal effects of policies or programs and the extent to which they are achieving their goals or objectives.

The Departments of Labor, Health and Human Services, and Energy, for instance, have assistant secretaries whose responsibilities include program evaluation. Bureaus within these and other departments often include policy- and program-evaluation staffs. Moreover, they enter into many contracts with private research organizations, university scholars, and others for the performance of evaluation studies. More attention appears to be given to evaluating social programs (welfare, education, health, and nutrition) than most other areas of governmental activity. This preference probably arises from the proliferation of social programs in recent decades, their substantial financial costs, the controversies that swirl around them, and the dislike of "welfare."

Systematic evaluation seeks information on the effects of a policy or program on the public need or problem at which it is directed. Utilizing particularly

the talents of social scientists (economists, political scientists, psychologists, and sociologists), it involves the specification of goals or objectives; the collection of information and data on program inputs, outputs, and consequences; and their rigorous analysis, preferably through the use of quantitative or statistical techniques.

Evaluation researchers use a variety of evaluation designs, which are often coupled with high-powered statistical and mathematical techniques. The three evaluation designs discussed here—the experimental design, the quasi-experimental design, and the before-and-after study—are basic.¹¹ An examination of them will convey a notion of the problems involved in setting up evaluation studies.

The *experimental design* is the classic method for evaluating a policy or program. Two comparable groups—an experimental, or treatment, group and a control group—are randomly selected from the target population. The experimental group receives treatment through a policy or program; the control group does not. Pretests and posttests of the two groups are used to determine whether changes, such as improved reading scores or lower incidences of a disease, have occurred in the two groups. If the performance of the experimental group is significantly better than that of the control group, the program is held to be effective. A high level of validity and generalizability is accorded the results of experiments. An example is in order.

Several years ago, the Delaware Department of Labor conducted a field experiment¹² to assess the effectiveness of various activities in helping “dislocated workers”—“persons who have lost long-term, stable jobs due to an increased international competition and/or changing technology.”¹³ The goals of the program for dislocated workers, which was funded through the Job Partnership Training Act, were to help increase workers’ earnings and reduce their need for unemployment compensation benefits. A coterie of 175 workers with comparable characteristics was identified; 65 were randomly assigned to the treatment group, with the other 110 becoming the control group. Members of the treatment group were provided counseling on job-search activities, assistance in locating openings, retraining, and other services. Both groups were monitored for a year. Comparison of their performance revealed that the control group did better than the treatment group in meeting the program’s goals. It was concluded that “the program did not appear to improve participants’ job prospects.”¹⁴

Use of the experimental design may not be possible because of costs, time, and ethical or other considerations. The *quasi-experiment*, then, may be a useful alternative. The process of random selection is not used. Rather, the treatment group is compared with another group (a “comparison group”) that is similar in many respects. Consequently, in the quasi-experiment there is greater likelihood that the performances of the two groups could be influenced by their internal characteristics rather than the program treatment. Quasi-experiments nonetheless are still seen as quite useful for many purposes.

A well-known quasi-experiment involves the Connecticut highway speed-control program.¹⁵ Following a record number of highway traffic fatalities,

the state initiated a crackdown on speeding. Initial data indicated that the enforcement program had significantly reduced fatalities. It was possible, however, that this resulted from other factors, such as differences in the weather or more safe cars in operation. To control for such possibilities, the Connecticut highway fatality rate per 100,000 people was compared with that of neighboring states where there had been no enhanced enforcement program. This showed that Connecticut’s fatality rate was lower than those of the other states, thus supporting the inference that the Connecticut crackdown had a positive effect in reducing traffic deaths.

The *before-and-after study* of a program compares the results of a program after a period of implementation with the conditions existing prior to its inception. Thus, one might compare the quality of water in a river before and after a pollution-control program was put into effect. Before-and-after studies often have low costs and take less time to conduct. A major drawback, however, is that the changes that occur are open to rival explanations. Improved water quality in our imaginary river could be due to increased flow, voluntary action by polluters, or an economic recession’s having caused reduced industrial activity. On the other hand, if a before-and-after study finds little change in the desired direction, then it is likely that the program is not having much effect.¹⁶ Given all of this, it is still possible for before-and-after studies to produce much information about a program that otherwise would be unavailable.

As this discussion indicates, systematic evaluation draws on experience in assessing the effects a policy or program has on the public need or problem at which it is directed. It permits at least tentative, informed responses to such questions as: Is this policy achieving its stated objectives? What are its costs and benefits? Who are its beneficiaries? What happened as a consequence of the policy that would not have happened in its absence? Consequently, systematic evaluation gives policy-makers and the general public, if they are interested, some notion about the actual impact of policy and provides discussions of policy with some grounding in reality. Evaluation findings can be used to modify current policies and programs and to help design others for the future.

Of course, evaluation studies can also be used for less laudable purposes. Professor Carol Weiss comments, “Program decision-makers may turn to evaluation to delay a decision; to justify and legitimate a decision already made; to extricate themselves from controversy about future directions by passing the buck; to vindicate the program in the eyes of its constituents, its funders, or the public; to satisfy conditions of a government or foundation grant through the ritual of evaluation.”¹⁷ In short, evaluators may be motivated by self-service as well as public service and by a desire to use analysis as ammunition for partisan or personal political purposes. Thus, the staff of the Federal Paperwork Commission was interested only in evaluations supportive of their goal of eliminating as much governmental regulation of business as possible, albeit under the guise of paperwork reduction.¹⁸

We now turn to some of the ways in which policy evaluation is handled in the national government.



Policy Evaluation Processes

Much policy evaluation is performed by nongovernmental actors. The communications media; university scholars and research centers; private research organizations such as the Brookings Institution, the Urban Institute, and the American Enterprise Institute; pressure groups; and public-interest organizations, Common Cause, the Audubon Society, and Ralph Nader and his "raiders" (a collection of public-interest activists) are examples; all make evaluations of public policies and programs that have greater or lesser effects on public officials. They also provide the general public with information, publicize policy success and failure, sometimes act as advocates for unpopular causes, and occasionally provide representation for those unrepresented in the policy process, such as the aged confined to negligently run nursing homes or exploited migratory farm workers.

Collectively, this aggregation of private evaluators of public policies has been designated by Professor Richard Hofferbert as the "policy evaluation industry."¹⁹ We will take a look here at three of its components.

Many university scholars engage in evaluation on their own initiative, as members of university research centers, while under contracts with government agencies, or to gain academic fame. Many universities have graduate schools of public policy whose faculty have a more "applied" or practical orientation and are inclined to undertake policy research. Some of the research of university scholars is published in such journals as *Evaluation and Program Planning*, *Evaluation Review*, and *Journal of Policy Analysis and Management*. For the most part, though, university scholars are more interested in making theoretical contributions to their academic disciplines than in getting involved in current policy debates.

A number of private research organizations ("think tanks") devote all or most of their time to performing evaluation studies.²⁰ The Manpower Development Research Corporation uses funds from national, state, and local governments and foundations to research social welfare and employment programs. The RAND Corporation operates mainly as a contract researcher producing policy reports for the Department of Defense. Research organizations of this sort often can act more independently and be more critical of agency actions than can internal agency personnel. "Of course, contract researchers often face pressures to follow the agency line, especially if they are highly dependent on a single agency."²¹

In their reporting of the news when public policies are involved, the media tend to focus on policies and government actions that go awry. Some sort of policy scandal is even more "newsworthy." The operative assumption seems to be that that sort of reporting will attract more viewers or readers than will information about a successful program or the routine operations of government. One of the television networks includes an occasional series titled "The Fleecing of America" as part of its evening news report. The focus is on governmental blunders and misadventures. Coverage of this nature contributes to the (incorrect) notion that "the government can't do anything well." More

positively, though, it may also help to focus public attention on some matters that need examination and correction. Television and radio programs that recount public-policy successes seem to be nonexistent.

Moving on, within the boundaries of the national government, policy evaluation is performed in numerous ways by officials and organizations. Sometimes these evaluations are rigorous and systematic; at other times, they are rather haphazard or sporadic. In some instances policy evaluation has become formalized, a regular component of the policy process; in others it remains essentially informal and unstructured. In the remainder of this section, we glance at a few forms of official policy evaluation—congressional oversight, Government Accountability Office (GAO) studies, the activities of presidential commissions, and "in house" evaluations handled within agencies.

Congressional Oversight

Although it is not specified in the Constitution, one of the primary functions of Congress is the supervision and evaluation of the administration and execution of laws and policies, which is commonly referred to as *congressional oversight*. Some, agreeing with the English political theorist John Stuart Mill, think that this is the most important function a legislature performs.²² Oversight, however, is not a separable, distinct activity; rather, it is an integral part of almost everything that members of Congress do, including gathering information, legislating, authorizing appropriations, and being of service to constituents. It may be intended either to control the actions of agencies, as when they sometimes are required to clear actions in advance with particular committees, or to evaluate agency actions, as when individual members or committees seek to determine whether administrators are complying with program objectives established by Congress. It is the evaluative aspect of oversight that is pertinent here.

Oversight may be performed by a number of techniques, including (1) casework, that is, intercession with agencies as a consequence of constituents' demands and requests; (2) committee hearings and investigations; (3) the appropriations process; (4) approval of presidential appointments; and (5) committee staff studies.²³ In the course of these activities, members of Congress reach conclusions about the efficiency, effectiveness, and impact of policies and programs—conclusions that can have profound consequences for the policy process. Congressional oversight is in essence more fragmented and disjointed than continuous and systematic. Bits and pieces of information, impressionistic judgments, and the members' intuition and values are blended to yield evaluation of policies and those who administer them. On the whole, however, members of Congress are more likely to be involved with initiation and adoption of policy than with evaluation, at least in any systematic sense.

The Congressional practice of enacting laws—for example, the Clean Air Act and the Head Start law—for determinate lengths of time builds evaluation into the legislative process. Periodically such laws have to be reauthorized by Congress, thereby creating opportunities to make changes in them on the basis of experience and available information. At the state level, sunset laws

have a similar effect. These laws provide that state agencies will be terminated unless they are periodically evaluated and reauthorized. That opportunities for evaluation are created does not, of course, guarantee that they will be well exploited.

Government Accountability Office

The GAO, usually regarded as an arm of Congress, has broad statutory authority to audit the operations and financial activities of federal agencies, to evaluate their programs, and to report its findings to Congress.²⁴ The agency, which has several thousand employees, has become increasingly involved with the evaluation of programs and now gives only a minor portion of its attention to financial auditing.

The Legislative Reorganization Act of 1970, which revamped the congressional committee system, also directed the GAO to "review and analyze the results of government programs and activities carried on under existing law, including the making of cost-benefit studies," and to make personnel available to assist congressional committees in handling similar activities. A subsequent statute authorized GAO to establish an Office of Program Review and Evaluation. Because of its expanded evaluation activities, the agency hires many more people trained in the social sciences than it once did.

Evaluation activities may be undertaken by the GAO on its own initiative, because of directives in legislation, at the request of congressional committees, or sometimes at the behest of individual members of Congress. In the course of a year, the GAO will produce several hundred evaluation studies, varying in length from several pages to a few hundred. The titles of three major studies issued by the GAO in mid-2004 indicate both the diversity and the evaluative thrust of its efforts: "Elections: Electronic Voting Offers Opportunities and Presents Challenges," "Gender Issues: Women's Participation in the Sciences Has Increased but Agencies Need to Do More to Ensure Compliance with Title IX," and "Maritime Security: Substantial Work Remains to Translate New Planning Requirements into Effective Port Security." GAO officials are also frequently called upon to testify before congressional committees.²⁵

Copies of GAO studies are sent to members of Congress and the affected agencies. The agencies are required by law to report to Congress and the Office of Management and Budget on actions taken in response to the GAO's recommendations or on why they did not act.

CASE STUDY

The GAO and Food Safety

A short case will help convey an understanding of the GAO evaluation process. Food safety is a continuing problem in the United States. The federal Centers for Disease Control and Prevention estimate that 5,000 people die in the United States annually from food-borne illnesses; millions more, many requiring hospitalization, suffer from food-related illnesses.

A dozen federal agencies operating under more than thirty laws govern the U.S. food supply. Primary among them are the Food and Drug Administration (FDA) in the Department of Health and Human Services; the Food Safety and Inspection Service (FSIS), the Agricultural Marketing Service (AMS), and the Grain Inspection, Packers and Stockyards Administration (GIPSA) in the Department of Agriculture; and the National Marine Fisheries Service (NMFS) in the Department of Commerce. The FDA regulates some 57,000 food-processing establishments and sets standards of quality and identity for food products. The FSIS supervises more than 6,000 plants engaged in slaughtering and processing meat and poultry products. The AMS has jurisdiction over hundreds of egg and egg-product facilities. The NMFS provides voluntary (fee for service) safety and sanitary inspections for approximately 2,500 foreign and domestic seafood firms.

In 1992, at the request of a House subcommittee chair, the GAO undertook an evaluation of the consistency, coordination, and effectiveness of the federal food-safety inspection program.²⁶ Data and information for the study were acquired by interviewing agency officials; by analyzing agencies' inspection policies, procedures, and records; and by meeting with industry and consumer groups.

The GAO reached several negative conclusions about the food-safety inspection system. Food products presenting similar health risks were treated differently by the various agencies, duplicative inspections occurred, and coordination among the agencies was inadequate. The agencies' effectiveness was limited by inconsistencies and illogical differences, based on statutory authority and funding, in their approaches to food safety. Thus, plants processing meat and poultry were inspected daily by the FSIS, but food-processing plants under FDA jurisdiction were inspected once every three to five years.

Further, the agencies drew some narrow jurisdictional lines. For one, the FSIS handled inspection of open-faced meat sandwiches made with one slice of bread whereas the traditional sandwich made with two slices of bread was the FDA's responsibility. For another, pizza with meat topping was inspected by the FSIS, pizza without meat topping by the FDA.

Finally, meat and poultry plants were required to register for inspection with the FSIS before they could operate and sell their products. In contrast, food-processing factories were under no such mandate and had to be tracked down by the FDA.

To correct these and other problems, the GAO recommended the creation of a single food-safety agency. Because this was not a good political possibility, the GAO said a "more realistic" alternative would be the appointment of a panel of experts to develop an inspection model based on public-health risks and strong enforcement authority. This would provide a rational basis for improving the effectiveness of the food-safety system. No action was taken on the GAO's recommendations.

Since 1992, the GAO has conducted dozens of studies relating to the food-safety system and calling for reform, especially for the creation of a