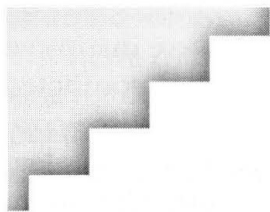


shortages. In other instances it showed little progress in achieving outcome goals.³⁴ In the case of the Environmental Protection Agency it was reported that “the annual performance measures established under GPRA are often selected on the basis of available data that focus primarily on outputs rather than environmental results for which credible data are often lacking.”³⁵



Problems in Policy Evaluation

The most useful form of policy evaluation for policy-makers and administrators, and for policy critics who want a factual basis for their positions, is a systematic evaluation that tries to determine cause-and-effect relationships and rigorously measures the results of policy. It is of course often impossible to measure quantitatively the effect of public policies, especially social policies, with any real precision. In this context, then, to “measure rigorously” is to seek to assess policy impacts as carefully and objectively as possible, using the best information available and making careful judgments. There is no reason to assume that “if it cannot be counted, it does not count.”

Determining whether a policy or program is doing what it is supposed to do, or doing something else, is not an easy, straightforward task, as some appear to assume. Snap judgments are easy to make but lack definitiveness. A variety of conditions raise obstacles or create problems for the effective accomplishment of policy evaluation. These include uncertainty over policy goals, difficulty in determining causality, diffuse policy impacts, and others, all of which are reviewed in this section.

Uncertainty over Policy Goals

When the goals of a policy are unclear, diffuse, or diverse, as they frequently are, determining the extent to which they have been attained becomes a difficult and frustrating task.³⁶ This situation is often a product of the policy adoption process. Because the support of a majority coalition is needed to secure adoption of a policy, it is usually necessary to appeal to persons and groups possessing differing interests and diverse values. To win their votes, commitments to the preferred policy goals of these various groups may be included in the legislation. The Model Cities Act, which was a major attempt to deal with urban problems, reflected this technique. Its goals included rebuilding slum and blighted areas; improving housing, income, and cultural opportunities; reducing crime and delinquency; lessening dependency on welfare; and maintaining historic landmarks. No priorities were assigned to the various goals, nor were their dimensions well specified. Model Cities evaluation research had to try to come to grips with the extent to which these diverse goals were being accomplished.

Determining the real goals of a program can be a difficult or conflictual task. Persons occupying different positions in the policy process, such as legislators and administrators, or national and state officials, or possessing differing ideological or philosophical perspectives, may define goals differently, act accord-

ingly, and reach differing conclusions about a program's accomplishments or success. Moreover, "because 'what you measure is what you get,' choosing the right goals to measure is essential."³⁷ Later in this chapter we will see how the multiple goals of the Head Start program, and measurement problems, have complicated its evaluation.

**Difficulty in
Determining
Causality**

Systematic evaluation requires that societal changes must be demonstrably caused by policy actions. The mere fact that when action A is taken condition B develops does not necessarily mean the presence of a cause-and-effect relationship. Other actions (or variables) may have been the actual causes of condition B. As we know, many common colds are "cured," not by ingesting medicines, applying ointments, or using nasal sprays, but by the human body's natural recuperative power.

Consider this example. Many states require periodic automobile safety inspections, in an attempt to reduce highway traffic accidents and fatalities. Research indicates that states with mandatory inspection laws do tend to have fewer traffic fatalities than do other states. Other factors, however, such as population density, weather conditions, and percentage of young drivers might in fact have more power in explaining the difference. Only if such conditions are controlled in the analysis, and if differences remain between states with and without inspections, can it be accurately stated that a policy of periodic automobile inspections reduces traffic deaths. In actuality, such laws do seem to have a modest beneficial effect.³⁸

To further illustrate the problem of determining causality, let us take the case of crime-control policies. The purpose, or at least one of the purposes, of these policies is deterring crime. *Deterrence* may be defined as the prevention of an action that can be said to have had a "realistic potential of actualization," that is, one that really could have happened.³⁹ (This assumption is required to avoid the kind of analysis that holds, for example, that consumption of alcoholic beverages prevents stomach worms, since no one has ever been afflicted with them after starting to drink.) The problem here is that not doing something is a sort of nonevent, or intangible act. Does a person's not committing burglary mean that he or she has been effectively deterred by policy from so acting? The answer, of course, first depends upon whether he or she was inclined to engage in burglary. If so, then was the person deterred by the possibility of detection and punishment, by other factors such as family influence, or by lack of opportunity? As this example indicates, the determination of causality between actions, especially in complex social and economic matters, frequently is a daunting task.

**Diffuse Policy
Impacts**

Policy actions may affect groups other than those at whom they are specifically directed. A welfare program may affect not only the poor but also others such as taxpayers, public officials, and low-income people who are

not receiving welfare benefits. The effects on these groups may be either symbolic or material. Taxpayers may grumble that their “hard-earned dollars are going to support those too lazy to work.” Some low-income working people may indeed decide to go on welfare rather than continue working at grubby, unpleasant jobs for low wages. So far as the poor who receive material benefits are concerned, how do benefits affect their initiative and self-reliance, family solidarity, or maintenance of social order? We should bear in mind that policies may have unstated intentions. Thus, an antipoverty program might have been covertly intended to help defuse the demands of black activists; or a program to control importing of beef may be intended to appease cattle growers politically, but not really do much to limit foreign competition.

The effects of some programs may be very broad and long-range in nature. Antitrust policy is an example. Originally intended to help maintain competition and prevent monopoly in the economy, how does one now evaluate its effectiveness? We can look at current enforcement activity and find that some mergers have been prevented and many price-fixing conspiracies have been prosecuted, but this record will tell us little about the extent of competition and monopoly in the economy generally. It would be pleasing to be able to determine that the economy is n percent more competitive than it would have been without antitrust policy. Because its goals are general and because measuring competition and monopoly is difficult, this determination just is not possible. Interestingly, after a century of antitrust action, we are still without agreed-upon definitions of monopoly and competition to guide policy action and evaluation. No wonder those assessing the effectiveness of antitrust policy sometimes come to sharply different conclusions.⁴⁰

Difficulties in Data Acquisition As implied in some previous comments, a shortage of accurate and relevant statistical data and other information may handicap the policy evaluator, particularly when one’s concern is with policy outcomes. Thus an econometric model may predict how a tax cut will affect economic activity, but suitable data to indicate its actual impacts on the economy are hard to come by. Again, think of the problems in securing the data needed to determine the effect on criminal law enforcement of a Supreme Court decision such as *Miranda v. Arizona*,⁴¹ which held that a confession obtained when a suspect had not been informed of his or her rights when taken into custody was inherently invalid. The members of the President’s Crime Commission in 1967 disagreed about its effect, the majority saying it was too early to determine results. A minority, however, held that, if fully implemented, “it could mean the virtual elimination of pretrial interrogation of suspects. . . . Few can doubt the adverse effect of *Miranda* upon the law enforcement process.”⁴² Absence of data does not necessarily hinder all evaluators.

The use of “*Miranda* cards” to inform suspects of their rights now has become standard police practice. A consensus exists among criminal-justice scholars and law-enforcement officers to the effect that this reform has had lit-

the adverse effect on law enforcement. Various field and quantitative studies support this view. Moreover, it is suggested that the Miranda rule has helped improve professionalism among the police.⁴³

For many social and economic programs, a question that typically arises is, "Did those who participated in programs subsequently fare better than comparable persons who did not?" Providing an answer preferably involves an experimental evaluation design utilizing a control group. The difficulty in devising a control (or comparison) group for a manpower program is summed up in this passage:

A strict comparison group in the laboratory sense of the physical sciences is virtually impossible, primarily because the behavior patterns of people are affected by so many external social, economic, and political factors. In fact, sometimes the legislation itself prevents a proper comparison group from being established. For example, the Work Incentive Program legislation required that all fathers must be enrolled in the WIN program within 30 days after receipt of aid for their children. Therefore, a comparison group of fathers with comparable attributes to those fathers enrolled in the program could not be established. Even if all the external factors of the economy could be controlled, it would still be impossible to replicate the social and political environment affecting any experimental or demonstration program. Thus, it is easy for a decision maker to discount the results of almost any evaluation study on the basis that it lacks the precision control group.⁴⁴

Because of problems such as those mentioned in the quotation, experimental designs frequently cannot be used. (This reason is apart from their often high dollar cost.) Second-best alternatives must then be utilized, such as a quasi-experimental design using a nonequivalent control group.⁴⁵

Official Resistance

Evaluating policy, whether it be called policy analysis, measurement of policy impact, or something else, involves reporting findings and making judgments on the merits of policy. This is true even if the evaluator is a university researcher who thinks that he or she is objectively pursuing knowledge. Agency and program officials will be alert to the possible political consequences of evaluation. If the results do not come out "right" from their perspective, or worse, if the results are negative and come to the attention of decision-makers, their program, influence, or careers may be thrown in jeopardy. Consequently, program officials may discourage or disparage evaluation studies, refuse access to data, or keep incomplete records.

Within agencies, evaluation studies are likely to be most strongly supported by higher-level officials, who must make decisions about the allocation of resources among programs and the continuation of given programs. They may, however, be reluctant to require evaluations, especially if their results may have a divisive effect within the agencies. Finally, we should note that organizations tend to resist change, and evaluation implies change. Organizational inertia may thus be an obstacle to evaluation, along with more overt forms of resistance.

A Limited Time Perspective

The time horizon of legislators and other elected officials often extends only as far as the next election. Consequently, they, and others who think like them, often expect quick results from governmental programs, even social and educational programs whose effects may take many years to fully appear. This being the case, short-run evaluations of program accomplishments may be unfavorable. A good example is the New Deal's resettlement program, which provided opportunities for land ownership to thousands of black sharecroppers in the South during the late 1930s and early 1940s. It was judged as a failure and just another New Deal boondoggle by contemporary critics. A decades-later evaluation of the program by policy analyst Lester Salamon concluded, however, that it had significant, positive, long-term effects, although not as an agricultural policy.⁴⁶ At modest cost, it did transform "a group of landless black tenants into a permanent landed middle class that ultimately emerged in the 1960s as the backbone of the civil-rights movement in the rural South." If the time dimension is ignored in evaluation studies, the results may be flawed and neglect important long-term effects. The pressure for rapid feedback concerning a policy can then create a dilemma for the evaluator.

Evaluation Lacks Influence

Once completed, an evaluation of a program may be ignored or attacked as inconclusive or unsound on various grounds. It may be alleged that the evaluation was poorly designed, the data used were inadequate, or the findings are inconclusive. Those strongly interested in a program, however, whether as administrators or beneficiaries, are unlikely to lose their affection for it merely because an evaluation study concluded that its costs are greater than its benefits. Moreover, there is also the possibility that the evaluation is flawed.

Governmental programs are not terminated solely as a consequence of an unfavorable systematic evaluation, although such evaluations did contribute to adoption of the Airline Deregulation Act (see further on). Of course, evaluations frequently lead to incremental changes or improvements in the design and administration of programs. That is the intent of many program evaluations done by the General Accounting Office, for instance, which, perhaps, is all that should be asked or expected of most evaluations.

**Policy Evaluation: The Use and Misuse of Cost-Benefit Analysis**

Cost-benefit analysis is a formal, quantitative evaluation technique that requires identifying the costs and benefits of either a proposed or actual policy and translating them into monetary values for purposes of comparison. It assumes that society will be made better off only by policies (or projects, or programs) whose benefits exceed their costs. Cost-benefit analysis has been most frequently used to evaluate proposed policies. Sometimes, though, it is employed to appraise existing

policies. Thus economist A. Myrick Freeman III used it to evaluate national air- and water-pollution-control policies. He found that the control of air pollution from stationary sources yielded benefits that were much greater than control costs. On the other hand, the costs of controlling industrial and municipal sources of water pollution were greater than the benefits realized.⁴⁷ In the following discussion, the focus will be on cost-benefit analysis primarily as a prospective evaluation technique.

The major steps in performing a cost-benefit analysis can readily be summarized.⁴⁸ First, one identifies all of the effects or consequences of a policy and categorizes them as costs or benefits for various groups. (Note that this requires establishing which groups are entitled to be considered in determining costs and benefits.) Both direct and indirect effects should be analyzed. Second, dollar values are placed on the various costs and benefits. This will be relatively easy for items that are customarily bought and sold in markets. For such matters as good health, the prolongation of human life, or scenic vistas, it will be much more difficult. Third, because some of the consequences of a policy will be current or short-term but others will occur many years hence, a discount rate is needed to equate the value of present and future effects. The basic assumption underlying the discount rate is that a dollar today is worth more than a dollar a decade or two from now. Inflation, for instance, may diminish the dollar's value, or purchasing power. Fourth, the costs and benefits, direct and indirect, current and future, of the policy are compared. If benefits exceed costs, the policy is acceptable; conversely, if costs exceed benefits, it should be rejected, or a better way of doing it should be found.

So presented, cost-benefit analysis appears as a reasonably clear-cut method for appraising policies. In actuality, however, there are significant problems involved in its application, a few of which are examined here.

Good data on the costs and benefits of a policy are frequently difficult to come by. How, for example, does one calculate the value of the health benefits of cleaner air. Or of the esthetic benefits of reducing haze in national parks? How are dollar values to be assigned to such matters? The value of land flooded (a cost) for a reservoir can readily be determined by reference to the value of nearby land. But what of the value of an ancestral home located there? The data and dollar values on which a cost-benefit analysis is based can be of tenuous and arguable nature.

It is, further, no easy task to identify the appropriate discount rate. It can be based on such criteria as the interest rate, the rate of inflation, or the opportunity costs of capital—that is, the rate of return that money would earn if devoted to private investment rather than public purposes. Despite its importance, there is no scientific way to decide on a discount rate. A low discount rate preserves the value of future benefits, whereas a high discount rate can sharply reduce their value. During the Reagan years, the Office of Management and Budget advocated a discount rate of 10 percent. This discount rate meant that the value of future benefits, such as lives prolonged two or three decades hence by reducing the incidence of cancer, would have very low value. This in turn increased the likelihood that a cost-benefit ratio would be unfavorable.

Cost-benefit analysis is based on the premise that efficiency is the primary, if not the only, value to be realized. Actions are evaluated on the basis of whether resources are used to improve the aggregate public good.⁴⁹ Little attention is accorded alternative or competing values—equity, human dignity, personal freedom, and equality, to name some. These are important to most people. The American system of criminal justice, for instance, is not very efficient because of our concern with equity and due process.

Another problem that arises in the course of many cost-benefit analyses is the need to place a value on a human life. Some take the position that life is priceless and that attempting to place a dollar value on life reduces it to just another commodity. In response it is argued that many policy decisions, such as industrial safety standards and highway speed limits, have an impact in terms of human lives. It is better to objectively take into account the value of life. Figure 7.1 displays some alternative ways for valuing human life.

FIGURE 7.1

The Valuation of Human Life

Several techniques for the valuation of human life for cost-benefit analyses have been developed.* Four are presented here.

1. *The human capital approach.* This is sometimes also called *discounted future earnings*. This technique, which appears free of moral sentiment, holds that a person's value depends upon what he or she can earn in the marketplace during a lifetime of work, discounted to the present. Some analysts would subtract a person's living expenses from earnings to arrive at a net value. The more one earns, the more one's life is worth.

2. *Willingness to pay.* This may be determined in a couple of ways. Wage rates in risky occupations may be compared with those in less risky occupations. Wage differentials and differentials in the magnitude of risk involved are used to calculate a life's value. For example, if annual wages were \$50 higher for a job that exposed someone to a 1 in 50,000 greater risk of death, that would yield a figure of \$2.5 million.

The contingent valuation variant involves using surveys to determine how much people are willing to pay to reduce health risks. Based on their responses, a dollar value for life is assigned.

3. *Court case settlements.* Awards made by judges and juries in product liability and medical malpractice cases for the loss of life could be used to construct the value of a life. It must be noted, though, that these awards greatly vary in amount and that many potentially actionable cases of negligent deaths are not litigated.

4. *Individual appraisal.* Assuming that people are the best judges of their own self-interest, they might be surveyed on how much they would be willing to pay to avoid death. Alternatively, a person could be queried about how much he or she would accept in payment for his or her death. Whether this technique would yield much usable information is doubtful.

* This listing draws on Kenneth J. Meier, "The Limits of Cost-Benefit Analysis," in Lloyd G. Nigro, ed., *Decision-Making in Public Administration* (New York: Marcel-Dekker, 1984), pp. 43–63; and Thomas O. McGarity, *Reinventing Rationality: The Role of Regulatory Analysis in the Federal Bureaucracy* (New York: Cambridge University Press, 1991), chap. 9.

Finally, let us note that cost-benefit analysis emphasizes the consequences for society as a whole. As we know, however, public policies distribute advantages and disadvantages, or costs and benefits. Those who pay the costs of policies often do not benefit from them, and vice versa. Put differently, policies have distributive consequences that are of importance. People may appropriately be more concerned with who benefits from industrial safety policies than whether their total costs exceed their benefits.

Problems such as those sketched here have not prevented cost-benefit analysis from being used as a tool in governmental decision-making for several decades. The Flood Control Act of 1936 specified that flood-control projects could be undertaken by the Army Corps of Engineers only "if the benefits to whomsoever they may accrue are in excess of the estimated costs." This standard must also be used for water projects handled by the Soil Conservation Service and has been voluntarily employed by the Bureau of Reclamation. In the 1960s cost-benefit analysis was first used in evaluating defense programs and then domestic programs as part of PPBS.

In the 1970s Presidents Ford and Carter directed executive-branch regulatory agencies to prepare "inflation impact statements" and "regulatory analyses," respectively, in developing some proposed regulations. These statements involved analyzing their expected economic consequences. The Carter administration made it clear, however, that although regulatory agencies should consider the burdens and gains of proposed regulations, a cost-benefit test was not to be used in appraising them.⁵⁰

A goal of the Reagan administration when it took office was to substantially reduce governmental regulation of private economic activity. People who were critical of the programs under their jurisdiction were appointed to regulatory positions. A second action involved issuing Executive Order 12291 in February 1981,⁵¹ which drew heavily upon the Carter administration's experience. The order required that proposed major regulations issued by executive-branch agencies (the independent regulatory commissions were exempt) must be accompanied by regulatory impact analyses assessing the potential benefits, costs, and net benefits of the regulations, including effects that could not be quantified in monetary terms, unless such calculations were prohibited by law. Some statutes ban use of cost-benefit analysis for the programs they establish.

Major regulations were defined as those likely to have an annual impact on the economy of \$100 million or more, to lead to major cost or price increases, or to have "significant adverse effects on competition, employment, investment, productivity, innovation, or the ability of U.S.-based enterprises to compete with foreign-based enterprises in domestic or export markets." The OMB was authorized to make the final determination of what was a major rule, to supervise the evaluation process, and to delay the issuance of proposed or final rules if it found the regulatory analyses were unsatisfactory.

Rules could be issued only if their estimated benefits exceeded their estimated costs. If a choice was available, the less costly alternative was to be selected. The burden of proof that this standard was met rested with the agency.

An action by the OMB holding up a rule could be appealed to the President's Task Force on Regulatory Relief, which was staffed by the OMB and comprised several executive officials under the leadership of Vice President George Bush. (The word "relief" in the task force's title indicates its orientation.) Although the task force was phased out in 1983, all of this planning was intended to ensure, among other things, that "Regulatory Action shall not be undertaken unless the potential benefits to society for the regulation outweigh the potential costs to society." Thus cost-benefit analysis was to be more than an analytical technique; it became a decision rule with a conservative bias.

The Reagan regulatory analysis program was a center of controversy. Critics contended that it was used improperly to reduce the extent of regulation and to delay the issuance of rules rather than to improve the quality of regulations by encouraging better analysis. The OMB was also accused of improperly interfering in the regulatory process by usurping authority vested in the regulatory agencies. The administration denied such accusations. In practice, though, administration officials demonstrated much more vigilance about the costs than the benefits of regulation in trying to reduce the burden of regulatory activity on businesses.

The George Bush administration continued the regulatory-analysis program and in time created the Council on Competitiveness, an interagency committee chaired by Vice President Dan Quayle, to work with the OMB in perpetuating the use of cost-benefit analysis.⁵² In the final two years of the George Bush administration, the Council on Competitiveness acted vigorously to represent the business community in the regulatory process and to reduce the number and strength of new regulations.⁵³ For the most part, it avoided publicity and sought to leave few "fingerprints."

The Clinton administration quickly abolished the Council on Competitiveness and later replaced Executive Order 12291 with its Executive Order 12866, entitled "Regulatory Planning and Review." The Clinton regulatory review program continued to make use of cost-benefit analysis for major rules. Regulatory review, however, was conducted more openly, and less stringently and intrusively than under the Reagan and George Bush administrations. Regulatory agencies experienced little to complain about.

The George W. Bush administration retained Clinton's executive order. The Office of Information and Regulatory Affairs, however, now came under the direction of a true believer in cost-benefit analysis. In six months OIRA rejected twenty-one proposed regulations, more than had been turned down by the Clinton administration in eight years.⁵⁴ This was not surprising given the strongly conservative, antiregulation stance of the George W. Bush administration. As this record indicates, presidential review of proposed regulations and the use of cost-benefit analysis have become regularized features of the regulatory process. How they are implemented will depend significantly upon the ideological leanings of a presidential administration.

Fairly used, cost-benefit analysis can contribute to rationality in the decision-making process by aiding in the identification and appraisal of alterna-

tives, helping to identify impacts or consequences, and otherwise developing information and insights that will assist persons in making reflective, well-considered decisions. A careful appraisal of the likely costs and benefits of a proposed action, and of the persons and groups upon whom they will fall, is certainly useful, regardless of whether all are converted into dollar figures, and without converting cost-benefit analysis into a decision rule.

Cost-benefit analysis, however, is open to manipulation to support the values and preferences of its users. In the instance of Executive Order 12291, because of the stoutly antiregulatory orientation of its implementers, it became a form of partisan political analysis dressed up as regulatory rationality.⁵⁵ Again, it is doubtful that the Army Corps of Engineers has ever been unable to undertake a rivers and harbors project that its officials really wanted to construct because a favorable cost-benefit analysis could not be contrived.

Policy evaluation, as our discussion indicates, is more than a technical or objective analytical process; it is also a political process. In the next section, a case study of the Head Start program illustrates how political factors can affect the conduct and results of an evaluation of a social program. The case also demonstrates that such evaluations, even when intended to be neutral or objective in form, become political because they can affect allocation of resources.

CASE STUDY

The Politics of Evaluation: Head Start

In January 1965, President Lyndon Johnson announced that a preschool program named Head Start would be initiated as part of the Community Action Program (CAP) authorized by the Economic Opportunity Act of 1964. The Head Start program, which was designed to help overcome the effects of poverty on the educational achievement of poor children, included early classroom education, nutritional benefits, parent counseling, and health services.

Initially, \$17 million in CAP funds was earmarked for summer 1965 to enable 100,000 children to participate in Head Start. The announcement of the program, however, produced requests for a much larger volume of funds from many localities. Officials in the Office of Economic Opportunity (OEO), who had jurisdiction over the program, decided to meet this demand. Ultimately, \$103 million was committed to provide places for 560,000 children. To say the least, the Head Start program was highly popular, undoubtedly because it directed attention to poor preschool children, who readily aroused the public's sympathy, and to the goal of equal opportunity.

Late in the summer of 1965, Head Start became a permanent part of the antipoverty program. According to President Johnson, Head Start had been "battle-tested" and "proven worthy." It was expanded to a full-year program. In fiscal year 1968, \$330 million was allocated to provide places for 473,000 children in summer programs and another 218,000 in full-year programs, making Head Start the largest component of the CAP. Essentially, Head Start was a multifaceted program for meeting the needs of poor children. More

than a traditional nursery school or kindergarten program, it was designed also to provide poor children with physical and mental health services and nutritious meals to improve their diet. Further, an effort was made to involve members of the local community in the operation of the program.

With this background, let us turn to evaluation of the program.⁵⁶ The OEO was among the agency leaders in efforts to evaluate social programs because of statutory requirements. Within the agency the task of evaluating its programs for overall effectiveness was assigned to the Office of Research, Plans, Programs, and Evaluations (RPP&E). Some early efforts had been made to evaluate the effectiveness of Head Start, mostly by Head Start officials and involving particular projects, but, by mid-1967, no solid evidence was available on overall program effectiveness. This lack was beginning to trouble OEO officials, the Bureau of the Budget, and some members of Congress. Consequently, the Evaluation Division of RPP&E, as part of a series of national evaluations of OEO programs, proposed an *ex post facto* study design for Head Start in which former Head Start children currently in the first, second, and third grades of school would be given a series of cognitive and affective tests. Their test scores would then be compared with those of a control group who had not been in the Head Start program. The Evaluation Division believed such a design would yield results more quickly than a longitudinal study that, although more desirable, would take longer to complete. (A longitudinal study examines the effect over a period of time of a program on a given group.)

Within OEO, Head Start officials opposed the proposed study on various grounds, including its design, the test instruments to be used, and the focus on only the educational aspect of the program to the neglect of its other goals—health, nutrition, and community involvement. The RPP&E evaluators acknowledged the multiplicity of Head Start goals but contended that cognitive improvement was its primary goal. They agreed with Head Start officials that there were risks in making a limited study, such as possibly misleading negative results, but insisted that the need for evaluative data necessitated taking the risks. In the wake of much internal debate, the OEO director decided the study should be made, and in June 1968, a contract was entered into with the Westinghouse Learning Corporation and Ohio University. The study was conducted in relative quiet, but hints of its negative findings began to surface as it neared completion.

Early in 1969, a White House staff official became aware of the Westinghouse study and requested information on it because the president was preparing an address on the Economic Opportunity Act that would include a discussion of Head Start. In response to the request, OEO officials reported the preliminary negative findings of the study. In his message to Congress on economic opportunity on February 19, 1969, President Nixon referred to the study, commenting that “the preliminary reports . . . confirm what many have feared: the long term effect of Head Start appears to be extremely weak.” He went on to say that “this must not discourage us” and spoke well of the program. Nonetheless, his speech raised substantial doubts about Head Start among many observers in the public arena.

The president's speech also touched off considerable pressure for release of the study's findings. The OEO officials were reluctant to do this because what had been delivered to them by Westinghouse was a preliminary draft, which was intended for use in deciding such matters as what additional statistical tests were needed and what data required reanalysis. From Congress, where hearings were being held on OEO legislation, claims were made that the study was being held back to protect Head Start and that the report was going to be rewritten. The pressure on the White House became sufficiently great that it directed OEO to make the study public by April 14. A major conclusion of the report was that the full-year Head Start program produced a statistically significant but absolutely slight improvement in participant children.

The release of the report set off a flood of criticism from Head Start proponents, including many academicians, directed at the methodological and conceptual validity of the report. A sympathetic article on the front page of *The New York Times* bore the headline "HEAD START REPORT HELD 'FULL OF HOLES.'" Much of the ensuing controversy focused on the statistical methods used in the study and involved a broad range of claims, charges, rebuttals, and denials. The proponents of Head Start seemed to fear that their program was being victimized by devious intent. This fear had several facets. One was that persons within OEO who favored Community Action over Head Start wanted a study that would spotlight Head Start's deficiencies. Another was that the administration was going to use the findings to justify a major cutback in Head Start. Finally, there was the fear that "enemies of the program" in Congress would use the negative results as an excuse for attacking it. Although there later appeared to have been little factual basis for these fears, they were real to the proponents of Head Start and contributed to the intensity of their assault on the evaluation study.

The methodological conflict which arose over the study focused on such standard items as sample size, validity of the control group, and appropriateness of the tests given the children. An examination of these matters would be too lengthy and too technical to include here. An assessment of the study by economist Walter Williams, however, provides a balanced view of the controversy:

In terms of its methodological and conceptual base, the study is a relatively good one. This in no way denies that many of the criticisms made of the study are valid. However, for the most part, they are the kinds of criticisms that can be made of most pieces of social science research conducted outside the laboratory, in a real-world setting, with all of the logistical and measurement problems that such studies entail. And these methodological flaws open the door to the more political issues. Thus, one needs not only to examine the methodological substance of the criticisms which have been made of the study, but also to understand the social concern which lies behind them as well. Head Start has elicited national sympathy and has had the support and involvement of the educational profession. It is understandable that so many should rush to the defense of such a popular and humane program. But how many of the concerns over the size of the sample, control-group equivalency, and the appropriateness of covariance analysis, for example, would have been registered if the study had found

positive differences in favor of Head Start? We imagine that this type of positive, but qualified assessment will fit any relatively good evaluation for some time to come. We have never seen a field evaluation of a social action program that could not be faulted legitimately by good methodologists, and we may never see one.⁵⁷

Interestingly, the findings of the Westinghouse study were as favorable to Head Start as were the earlier evaluations of specific projects made by Head Start officials. These, too, showed that the program had limited lasting effects on the children. What the Westinghouse study, and the controversy over it, did was to inject these findings into the public arena and expand the scope of the conflict over them.

Despite the essentially negative evaluations of its accomplishments, the Westinghouse report recommended that Head Start be preserved and improved, at least partly on the ground that "something must be tried here and now to help the many children of poverty who may never be helped again." Head Start was, and is, a politically popular program. Congress and the executive have generally been favorably disposed toward the program, and it has suffered little of the criticism directed at other aspects of the antipoverty program. Children are a potent symbol in policy conflicts.

Ten years after the Westinghouse study was made public, the findings of another group of researchers on the long-term effects of Head Start were published by the Department of Health, Education, and Welfare. Based on a series of longitudinal studies, this study concluded that Head Start had significant, long-lasting social and educational benefits for its participants. Thus children who had been in the program had much less need for remedial classes, were less likely to be retained in grade, and were half as likely to drop out of high school as were adolescents of comparable age who had not been in the program.⁵⁸ As a consequence, Head Start was now hailed as a success by the communications media. Why the substantial difference in findings by the two evaluations? The explanation rests primarily with the different methodological approaches. The Westinghouse study, using an experimental design, focused on short-run effects, especially as measured by intelligence-test scores. The second study focused on long-range effects.

In 1981, Head Start was designated part of President Reagan's "social safety net," which provided assistance to the "truly needy," and thus was not tagged for cutbacks in funding, as were several other programs that provided aid to poor people. In 1988 approximately 450,000 children were enrolled in Head Start, which now operated year-round, at a cost of \$1.2 billion. Only about a quarter of the eligible children were actually enrolled in the program, however. Head Start continued to expand under the Bush and Clinton administrations. The program's appropriation for fiscal year 1998 was \$4 billion, which provided funding for more than 830,000 enrollees.

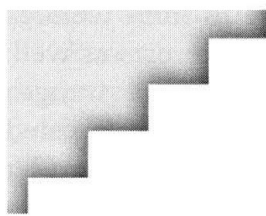
Research studies on the benefits of Head Start and early childhood education have continued to yield inconclusive findings. Children who go through Head Start are found to have improved cognitive abilities, greater self-esteem,

and improved social skills. On the other hand, various studies report that gains in academic achievement are not lasting. After a few years, when Head Start children are compared with non-Head Start children, the educational gains fade away.⁵⁹

A major evaluation of Head Start published in the *American Economic Review* in 1995 illustrates these mixed findings.⁶⁰ Using longitudinal data for a sample of nearly five thousand children, the evaluators examined the impact of Head Start on cognitive achievement, school performance (whether a child repeated one or more grades), utilization of preventive medical care, and health and nutritional status. Children who had been enrolled in Head Start were compared with their siblings who either had been enrolled in other preschool programs or had had no preschool experiences.

The evaluation found that Head Start had positive and persistent effects on the cognitive achievement and school performance of white children. In contrast, although there were positive effects on the cognitive achievement of African-American children, these effects soon disappeared. No positive effects were found on the school performance of African-American children. For both white and African-American children, Head Start had a positive effect on preventive health care, as measured by measles immunization rates. For neither did it have an impact on health and nutritional status, as measured by conformity with national height-for-age norms.

Whatever the results of the evaluation studies, Head Start has been a politically popular program, usually drawing bipartisan support from Congress and the executive. It is perceived as a way to provide educational and social services to those who are really in need—children in economically disadvantaged families. Most Head Start families have annual incomes of less than \$15,000. Perhaps, too, as many believe, it will lead to future reduced expenditures for other programs—welfare, juvenile delinquency, and criminal justice. In 2002, Head Start had a budget of \$6.5 billion and enrolled more than 900,000 children. ■



Policy Termination

As noted in the preceding section, the evaluation and appraisal of a policy, dissatisfaction with its costs and consequences, and the development and expansion of political opposition may produce a variety of responses to it, including termination. Policies are only one set of targets for termination. Others are programs (e.g., the rural abandoned-mine program), projects (e.g., the cross-Florida barge canal), and organizations. More than half of the states, for example, have enacted sunset laws, which require the legislature to renew periodically the authorization for administrative agencies. If this is not done, agencies are automatically abolished.

In this section the focus is on policies. Most of us can readily identify a number of government policies that we regard as wasteful, unnecessary, or inappropriate because they offend our ideological inclinations. Others, however,