Federal evaluations in education contributed in numerous and important ways to changes in law, regulation, and management. Information about implementation contributed most often, but information about outcomes, cost, and federal administration also made contributions. The findings of the present study are surprising in light of previous research that concluded evaluations do not often contribute to decisions.

CONTRIBUTIONS OF EVALUATION TO EDUCATION PROGRAMS AND POLICY

LAURA C. LEVITON
University of Pittsburgh
ROBERT F. BORUCH
Northwestern University

Truism of evaluation is that studies seldom contribute directly to identifiable decisions (Patton et al., 1977). The claim appears in several texts advocating reforms in evaluation. In Guba and Lincoln (1981) for example, it is averred that the failure to use evaluation findings has "almost assumed the proportions of a national scandal" (p.ix). In the "Ninety-Five Theses" of Cronbach et al., (1980), we find the assertion that "everyone agrees that evaluation is not rendering the service it should" (p. 3). Dunn et al., (1981) base their suggestions for reform on the presumption that evaluations are peripheral because they are seldom utilized.

The truism deserves a second look because the research on which it is based is almost a decade old, was scanty to begin with, and faced serious

AUTHORS' NOTE: *This research was supported by contract 300-79-0467 from the Department of Education to Northwestern University. We wish to thank John Evans, Janice Anderson, and the many people who patiently responded to our interviews. The work stems from a report to Congress and the Department of Education on the evaluation of federally supported education programs at the national, state, and local levels of government (Boruch and Cordray, 1980).*

EVALUATION REVIEW, Vol. 7 No. 5, October 1983 563-598

obstacles to accurately assessing use. While evaluators accept that the linkage to program change may not be straightforward, they also tend to believe that this linkage is mostly absent. Just as serious, Congressional aides may share the belief (Saxe and Koretz, 1982) and cut budgets accordingly.

Reform of evaluation may indeed be warranted on a variety of grounds, including the failure to use findings. However, reformers must do more than reassert the claim of nonuse. They must show that, for a given audience and setting, the level of influence is in fact low. They must also base their prescriptions on documented examples of evaluations that were used in important ways. Some examples exist, but more are needed to understand the variety of possible contributions.

The present study addresses empirically the issue of use and nonuse. It stems from a Congressionally mandated appraisal of educational evaluation at the national, state, and local levels of government (Boruch and Cordray, 1980). Our findings challenge the truism of nonuse for the federal level in education. Evaluations at this level were found to make frequent and important contributions to decisions. They were not the only input to those decisions, of course.

Some distinctions are important to avoid the confusion of earlier studies. We distinguish between "use" and "impact," defining the former as serious consideration of findings, which may or may not relate to decisions; and the latter as actual changes in programs resulting from use (Leviton and Hughes, 1981). The evidence that is brought to bear then concerns specific links between evaluation and serious consideration, decisions, and impact. Both oral testimony and paper trails were used as evidence and, moreover, we demanded corroboration of these trails from more than one source. It is possible that trails may end with serious consideration that enlightens decisionmakers and contributes indirectly (Weiss, 1977). Trails may end with decisions but no impact, because impact takes time, and change in programs involves decisions by many people (Pressman and Wildavsky, 1973). In other cases, trails may lead to amendments, regulations, and management changes.

METHOD

SAMPLE OF STUDIES

We drew two samples in this study. The first was from the evaluation unit of the Office of Education (OE). We selected fourteen evaluations

from a total of 31 completed studies described in the "Highlights" section of OE's Annual Report on evaluation for 1978 and 1979. Excluded from the definition of evaluative studies were reports on finance, management of technical assistance, and economic projections. Although we examined slightly less than half the evaluations, we cannot claim our sample is representative. We may have selected reports that were utilized more than the ones we did not study. However, even these were often used, according to descriptions (not verified by us) in the OE *Annual Reports.*

The second sample consisted of seven studies sponsored by a variety of federal agencies dealing with education. These studies were visible and widely discussed, according to those we interviewed. They were readily accessible to us and seemed a good place to find contributions. Both samples were purposive and, we believe, biased in favor of discovering contributions. Table 1 lists the studies examined, types of information covered, their general area of education policy, their sponsor, and the firm conducting the study.

Several kinds of information from these evaluations could contribute to programs and policy. Implementation findings indicate how local and state agencies delivered education programs. Outcome information consists of causal inferences about effects of programs on students and schools. Information about federal administration reveals problem areas of management, or addresses questions of entitlement and resource allocation. The cost and cost-effectiveness of programs constitutes another category. Frequently, several kinds of findings were involved in a single evaluation.[1] Sixteen of the studies, or 65%, presented implementation findings. Eleven of the studies, or 52%, presented program outcomes. Four studies, or 19%, presented information about federal administration, while almost all made some reference to costs.

DOCUMENTATION OF CONTRIBUTIONS

For each case study of a report, we developed a pool of informants and relevant documents by using a snowball sampling technique described by Rich (1979). The pool usually started with telephone interviews of the agency project officer and of the contractor who carried out the research. Sometimes, other respondents and relevant documents, such as Congressional reports, were available at the beginning of the case study. All respondents then named others who might be knowledgeable about use, and identified supporting documents.

TABLE 1
Evaluations Examined for Their Contributions

OE Sample	*Topic and Type of Information Provided*	*Policy Area*	*Contractor*[a]
1979	Services to Neglected and Delinquent Youth (implementation)	ESEA, Title I (Compensatory education)	Systems Development Corporation
	Sustaining Effects of Compensatory Education (implementation)	ESEA, Title I (Compensatory education)	Systems Development Corporation
	Magnet Schools (implementation, outcome)	ESAA (desegregation)	Abt Associates
	Survey of Educational TV Viewership (implementation)	ESAA (desegregation)	Applied Management Sciences
	Sex Equity in Programs (implementation)	Vocational Education	American Institutes for Research
	Campus-Based Aid and Basic Grant Programs (implementation)	Higher Education	Applied Management Sciences
	Project Implementation Packages (implementation, outcome)	ESEA, Title IV-C (innovations)	American Institutes for Research
1978	Aid to Non-Profit Organizations (implementation, outcome)	ESAA (desegregation)	RAND Corporation
	Facilities Survey (implementation)	Vocational Education	Westat
	Exemplary Programs (outcome)	Career Education	American Institutes for Research
	Upward Bound (outcome, implementation, Federal administration)	Higher Education	Research Triangle Institute
	Follow Through Planned Variations Study (outcome)	Economic Opportunity Act (Compensatory education)	Abt Associates
	Bilingual Education (outcome, implementation)	ESEA, Title VII	American Institutes for Research
	Survey of Programs (implementation)	Indian Education	Communications Technology Corp.

TABLE 1 (Continued)

Sector-Wide Sample Sponsor[b]	*Topic and Type of Information Provided*	*Policy Area*	*Contractor*[a]
ASPE	Fund for the Improvement of Postsecondary Education (outcome)	Higher Education	NTS Research Corporation
ASPE	Follow Through Exploratory Evaluation (Federal administration)	Economic Opportunity Act (Compensatory education)	in-house
ASPE	Impact Aid (Federal administration)	ESEA, Aid to Schools in Federally Impacted Areas	in-house
OE, 1977	Federal Programs Supporting Educational Change (implementation, outcome)	ESEA, Several Titles	RAND Corporation
NIE	Compensatory Education Study (implementation, Federal administration, outcome)	ESEA, Title I (compensatory education)	in house with many sub-contractors[a]
NIE	Achievement Testing (implementation)	ESEA, Title I (compensatory education)	SRI International
ACYF	National Day Care Study (implementation, outcome)	HEW Day Care (interagency)	Abt Associates

a. References can be found in Leviton and Boruch, 1980.

b. Abbreviations indicate the following organizations in the Department of Health, Education and Welfare:

ASPE — Assistant Secretary for Planning and Evaluation
OE — Office of Education
NIE — National Institute of Education
ACYF — Administration for Children, Youth and Families

To the extent possible, all information was verified from additional sources independent of the snowball sample. For example, Congressional staffers were asked about uses of research information in legislation, and independent observers supplied background information on the research reports.

For the purposes of our study, contributions had to be verified by more than one source. These sources could be two or more independent respondents, or documents supporting the contentions of a single respondent. However, statements by the project officer and the contractor or grantee were not counted as corroborating each other, because they had too similar an interest in showing the study's contributions. If, on the other hand, the contractor or project officer said that a third individual, such as a program manager, had used the information, the third individual's description of the use of the information was viewed as corroboration.

Unconfirmed Instances

Some contributions could not be confirmed easily from more than one source. Conceptual uses provide an example. If a single OE staff person said that findings helped him or her to outline legislative issues for the administration, this private use might not be confirmed by others, and the relevant memos might have been long gone. Rather than eliminate unconfirmed instances, we present them separately, in order to preserve information that has, admittedly a lower standard of evidence. These instances constituted 13% of the total number of contributions discovered.

CONTENT ANALYSIS OF CASE STUDIES

The contributions of these evaluations are described at length in Leviton and Boruch (1980). In order to provide a numerical summary, we characterized the types of use and impact involved, as well as the kinds of findings that contributed in each instance.

Types of Contributions

We classified contributions into those having impact on programs or policy, those that influenced decisions without achieving impact (at the

time of our investigation), and those that involved serious consideration but no decision that we could verify. Serious consideration might involve either the use of findings to persuade others to a point of view, as when lobbyists present their agenda and support it with findings; or conceptual use, as when an official tells us that evaluation revealed new problems or new solutions in a program. Evaluations contributing to impact have also contributed to decisions; evaluations contributing to decisions have also been seriously considered. The paper trail or oral history had to verify the linkages between these levels in order for the instance to be classified at the highest level.

Decision Contexts

The contributions we located occurred in specific decision contexts. Thus, in legislation, discussions of the implications of findings by Congressional staff, lobbyists, or education officials would be termed a use of information, but one that does not go beyond serious consideration. If people said that the findings were used in drafting a bill, or if the rationale for the bill cited findings, the instance was classified as use in decision-making. An actual change in legislation, in which independent parties claimed findings had been used, was classified as contribution to impact. Generally, a finding might partially motivate decisions or impact, or it might influence the form of decisions or impact.

Similarly, evaluations could contribute to consideration of the budget, to budget proposals (decisions), or to actual appropriations and authorizations (impact). Findings could contribute to discussions of regulations, to drafting them (decisions), or to final regulations (impact). Findings could be used in thinking about management, in decisions and orders, or in observable changes in management (impact). In similar ways, findings could contribute to states' and localities' decisionmaking. Reseach and development (R&D) activities were found to involve government analysis of data for other purposes, as well as for funding of new research projects. This category was classified as "decisionmaking" in that no impact on a program was evident in the short run. However, people could also "seriously consider" new R&D and how it should be carried out, as in books on evaluation methods.

One kind of serious consideration did not fit nicely into these decision contexts: The use of findings in editorials in the media. Findings were seriously considered by influential opinionmakers, but did not relate to specific decision contexts. These were the only such cases in the sample.

Unit of Analysis

We chose the individual contribution of a finding as the unit of analysis. This could be a qualitatively distinct impact, or a distinct decision. "Distinctiveness" was judged by the content of the contribution. Thus two decisions are counted in the following example from the RAND Study, "Federal Programs Supporting Educational Change":

> [One Administration Staff member] mentioned that her office was disturbed at the RAND finding that innovative projects were isolated from the rest of the school. They therefore inserted language in the Administration's proposal that local districts must make a commitment to spend some money on the project themselves, over time, while the federal share would decrease. [This office] also inserted language in the proposal that projects should be integrated into the rest of the school [Leviton and Boruch, 1980].

The decision on funding is different from the decision on integration. Alternative units could legitimately have been used; however most of them lose information.[2]

Because we chose qualitatively distinct contributions for analysis, and because agreement was needed about what constituted a contribution, inter-rater reliability was at issue. After training in classifying three case studies, two raters reached relatively good agreement on the remainder: a kappa of .84 in rating contributions, and .75 for types of information.

RESULTS

We first present our content analysis, to reveal the extent of contributions by types of information in different decision contexts. However, this numerical summary is indeed shallow unless the reader understands the nature of these contributions in their policy contexts. Thus, we highlight some of these contirbutions and suggest reason for them. Space does not permit a full presentation, which may be found in Leviton and Boruch (1980).

CONTENT ANALYSIS

We located 156 distinct contributions of evaluation. When unconfirmed instances are included, the total rises to 180. Our verification trail

led to impact 68 times (76 verified and unverified), to decisions only, 61 times (67 verified and unverified) and to "serious consideration" only, 27 times (37 verified and unverified). Because one can never be sure one has located all contributrions in a retrospective study, all numbers presented here should only suggest an emphasis given to types of information and types of contribution. The purposive sample may be biased in favor of detecting contributions; the retrospective method is biased against detection. We cannot even guess the direction of possible bias in favor of one kind of contribution over another.

Table 2 reveals these emphases. Looking at the types of contributions, it is clear that impacts are relatively more frequent than we might expect, given the difficulty of obtaining impact, and "consideration only" is surprisingly less frequent, given the evidence of past studies (Patton, et al., 1977; Weiss and Bucuvalas, 1977). In all likelihood, this retrospective method was biased toward reporting impacts and decisions, which people remember, and which are more frequently public acts. Nevertheless, the evidence is extremely revealing.

Implementation information contributed most frequently in this set of studies. Federal administration information contributed fairly often, especially when one remembers the small number of reports containing this information. Outcome information did contribute, and at high levels, but not as much as one might expect given its emphasis in the reports we studied. Cost and cost-effectiveness information did not contribute often in this set of studies (it was not a major focus for most), but contributed to impact in all three cases.

Table 3 presents contributions by the kind of information contributing, and the decision context of the contribution. We were able to locate primarily federal-level contributions of these federal-level reports. This may represent a bias in the method of sampling respondents, or it may be that findings were not used by state and local audiences. Our federal-level informants did not name many contacts at these levels.

Findings contributed most often to the form of legislation, both proposed and passed into law. Moreover, the implementation information, and to a lesser extent, federal administration information, contributed most often to legislation activities. Findings rarely contributed to budget allocations—but decisions, in the form of budget proposals to Congress, did incorporate findings relatively often, primarily to argue for increases or decreases. Again, implementation information contributed most often.

Contributions relating to the drafting of regulations may, with time, have become impacts as final regulations were approved.[3] Twelve of

TABLE 2
Contributions of Evaluation by Type of Information and Level of Contribution

Contributions[a]	*Type of Information*						
	Implementation	Outcomes	Implementation and Outcomes	Federal Administration	Federal Administration and Implementation	Cost and Cost-Effectiveness	Total
Total Contributions	67 (84)	22 (24)	25 (29)	29	10 (11)	3	156 (180)
Total Impact	28 (33)	9 (11)	15 (16)	10	3	3	68 (76)
Total Decisions only	29 (34)	4	5	16	7 (8)	0	61 (67)
Total "Consideration" only	10 (17)	9	5 (8)	3	0	0	27 (37)

Type of Information	*Proportion of Studies Presenting Information*[b]	*Proportion of Contributions by Information*[b]
Implementation	.76	.65 (.69)
Outcome	.52	.30 (.29)
Federal administration	.19	.25 (.22)
Cost effectiveness[c]	.10	.02 (.02)

a. Verified plus unverified instances appear in parentheses.
b. Proportions do not add to 1.0 because of overlap in content of studies and contributions.
c. Costs were mentioned in almost all reports.

these decisions were Congressional orders to the effect that federal education administrators draft regulations. These orders involved implementation and federal administration information. The officials writing the regulations tended to use implementation and the combination of implementation and outcome information.

TABLE 3
Contributions of Evaluation by Type of Information and Decision Context

Decision Context	*Type of Information*						
	Implementation	Outcomes	Implementation and Outcomes	Federal Administration	Federal Administration and Implementation	Cost and Cost-Effectiveness	Total
Legislative[a]							
Impact (Law)	21 (25)	1	4	8	3	1	38 (42)
Decisions (proposals)	8 (9)	0	0	6	1 (2)	0	15 (17)
"Consideration"	9 (10)	2	1	1	0	0	13 (14)
Budget							
Impact (authorization, appropriations)	1	2 (4)	1	0	0	0	4 (6)
Decisions (proposals)	5 (8)	0	0	0	2	0	7 (10)
"Consideration"	0	0	0	1	0	0	1
Regulations							
Impact (passage)	3	0	5	0	0	2	10
Decisions (drafting)	4 (5)	0	1	4	4	0	13 (14)
Management							
Impact	0 (1)	6	2	2	0	0	10 (11)
Decisions	5	4	1	2	0	0	12
"Consideration"	1 (2)	0	0 (2)	1	0	0	2 (5)
State							
Impact	0	0	3 (4)	0	0	0	3 (4)
Decisions	0	0	1	0	0	0	1
"Consideration"	0 (5)	0	0 (1)	0	0	0	0 (6)
Local							
Impact	3	0	0	0	0	0	3
Research and Development[b]							
Decisions	7	0	2	4	0	0	13
"Consideration"	0	3	0	0	0	0	3
Other "Consideration"							
Media Editorials	0	4	4	0	0	0	8

a. Verified plus unverified instances appear in parentheses.
b. Two mandated studies were included under legislative impact.

In the instances involving management, pure outcome information was used in ten exceptional cases, involving the Joint Dissemination Review Panel (JDRP). This panel approves funding for adoption and dissemination of educational innovations, based on evidence of effectiveness. Otherwise, managers did not use outcome information alone. This is especially surprising in the case of decisions to fund further research (R&D). The three instances of "consideration" of pure outcome data involve knowledge gained about how to run an evaluation. In most other instances, R&D was funded to solve problems revealed by implementation and federal administration information.

The National Institute of Education (NIE) Compensatory Education Study accounted for 43 of the contributions we discovered. The study was an "outlier" in terms of contributions, and unusual in other respects as well. To examine whether it contributed differently from other evaluations, we contrast it with others in Table 4. The NIE study contributed in 26 of the 38 instances of legislation, and in the 12 instances in which Congress ordered new regulations to be written. Congress made heavy use of NIE's information about implementation and federal administration, especially in legislation. The legislative contributions of other studies involved implementation most often as well, but less than half of these achieved impact. Although federal administration information from other studies was used, it did not achieve impact in legislation or regulation.

In Table 5, we present the contributions made by each of the studies. The median number of impacts was two, the median for decisions only was one, and the median for serious consideration was one (whether one looks at verified instances only, or verified plus unverified instances). Table 5 also reveals that our sector-wide sample of seven studies accounts for more than half of the total contributions discovered. The NIE Study is largely responsible for this; however, the study, "Federal Programs Supporting Educational Change," ranks second in the number of contributions.

HIGHLIGHTS OF CONTRIBUTIONS

Title I: NIE Compensatory Education Study

This study was a synthesis of 35 separate research projects, involving alternative eligibility criteria, services and their effects on students, and

TABLE 4
Contrast of NIE Compensatory Education Study with All Others

Contributions[a]	*Information*			
	Implementation	*Federal Administration*	*Implementation & Federal Admin.*	*Other*
Legislative:				
NIE	14	8	3	1
All Others	7 (11)	0	0	5
Legislative Bills:				
NIE	0	1	0	0
All others	8 (9)	5	1 (2)	0
Regulation:				
NIE	0	0	0	0
All Others	3	0	0	7
Regulation—Writing:				
NIE	4	4	4	0
All Others	0 (1)	0	0	1
Other Contributions:				
NIE	1	1	0	1
All Others	30 (41)	10	2	35 (41)

a. Verified plus unverified instances appear in parentheses.

management of the program at federal, state, and local levels. Readers interested in the details of the study should consult Hill (1980). Leviton and Boruch (in press) describe the various contributions of the study. Hearings on Title I took their form directly from the six topics of the reports. In addition, every respondent felt the contributions were important, and the Senate and House reports for the legislation specifically thank the staff and commend their highly useful work, which it "consequently relied on . . . in formulating Amendments to Title I " (House Report, p. 5.).

The information on implementation largely concerned troublesome variations in the state and local management of Title I programs. NIE

TABLE 5
Contributions Made by Each Study

OE Sample	*Contributions*[a]			
Topic	*Consideration, Decisions, and Impact*	*Consideration and Decisions Only*	*Consideration Only*	*Total*
Services to Neglected and Delinquent Youth	4	1	1 (2)	6 (7)
Sustaining Effects of Compensatory Education	0 (1)	5 (6)	3	8 (10)
Magnet Schools	3	2	2 (4)	7 (9)
Survey of ESAA TV Viewership	1	0 (1)	0	1 (2)
Sex Equity in Vocational Education Programs	0	0	0 (1)	0 (1)
Campus Based Aid and Basic Programs	0	6	0	6
Project Implementation Packages	3	1	1	5
Aid to Non-Profit Organizations	5	0	0	5
Vocational Education Facilities Survey	0	4 (6)	5 (10)	9 (16)
Exemplary Programs in Career Education	6	4	0	10
Upward Bound	0 (2)	1 (2)	0	1 (4)
Follow Through Planned Variations Study	0	0	7	7
Bilingual Education	5 (9)	4	1	10 (14)
Survey of Indian Education Programs	0	1 (2)	0	1 (2)
Total OE Contributions	27 (34)	29 (35)	20 (29)	76 (98)

TABLE 5 (Continued)

OE Sample	*Contributions*[a]			
Topic	*Consideration, Decisions, and Impact*	*Consideration and Decisions Only*	*Consideration Only*	*Total*
Sector-wide Sample				
Fund for the Improvement of Postsecondary Education	2	0	0	2
Follow Through Exploratory Evaluation	2	6	0	8
Impact Aid	0	5	1	6
Federal Programs Supporting Educational Change	4 (5)	5	3 (4)	12 (14)
NIE Compensatory Education Study	27	14	2	43
Achievement Testing	2	1	0	3
National Day Care Study	4	1	1	6
Total Sector-side Sample Contributions	41 (42)	32	7 (8)	80 (82)
Total Contributions	68 (76)	61 (67)	27 (37)	156 (180)

a. Verified plus unverified instances appear in parentheses.

attributed many of these problems to lack of clarity in law, regulation, and enforcement by OE—federal administration information. Congress clarified the law, borrowing from model legislation generated by NIE, and directed OE to clarify regulations. Examples include the methods for allocating federal and local compensatory education funds, legal models for programs in elementary and high school, and clarification of the states' role in monitoring compliance. In addition, outcome information was used both in changing Congress' thinking toward a more positive view of what the program could accomplish, and in arguments increasing the budget authorization.

The Study possessed many of the features claimed to be important in a usable evaluation: communication with users, asking relevant questions through careful planning, consultation with many stakeholders, and timeliness of results. In addition, most of the contributions related to changes in program administration. As a major source of federal education money, Title I was too politically impacted for Congress to consider drastic changes in the pre-Reagan era.

Title I: The Sustaining Effects Study

A preliminary report from this longitudinal study examined eligibility of students. The report and similar findings from the NIE study fueled a debate in Congress over the criteria of eligibility for Title I services. Because the Sustaining Effects Study showed only 39% of poor children were being served, the committee chairman could argue that funds should continue to be targeted to the poor. Because only 40% of low-achieving students were served, the ranking minority member could argue that funds should be targeted to this group regardless of income. The debate was not resolved, and neither side was happy with a study that could provide ammunition for both. Our view, of course, is that it did inform the debate—although it went no further than "serious consideration" in support of opposing policy positions.

The study was used, however, in federal management of the program; for example, in renewed efforts to better target services to appropriate children. The study took a form more useful to managers, who must rely on numbers served in program operations, than to Congress.

Title I: Achievement Testing

This study recommended testing students in the fall, to encourage uniformity of reporting. Scores tend to decline over the summer months, so that spring testing can inflate estimates of the effect of Title I programs. Congress amended the law, requiring that evaluation of Title I programs be based on achievement testing "over at least a twelve-month period in order to determine whether school year programs have sustaining effects over the summer." In addition, the law was changed to allow funding of summer bridge programs. The origins of this change lie

in an earlier bill that also cited this study. In this example we see how contributions of evaluation may take more than one legislative cycle.

Title I: Services to Neglected and Delinquent Youth

A major finding of this evaluation was that Title I served only half of the eligible students in state institutions for neglected and delinquent children. Moreover, Title I students received less time in instruction than non-Title I students. Congress used these findings in amending Title I, to insure that students would get more instructional time. The program manager gave this study his full cooperation, because he wanted policy makers to become aware of the problems and resolve them. The problems were beginning to be addressed as a result of the study. An exploratory evaluation was contracted to further describe the problems of state noncompliance. State and local educators took the study back to their institutions to argue for improvements. We were able to trace one Director of Education at a school for delinquent boys who used the study to convince the school to change the curriculum.

Desegregation: Aid to Non-Profit Organizations

These organizations received federal funds to facilitate school desegregation, but the evaluation found that they did not do this well. Congress used the study in converting the funding for these organizations from a state apportionment to competitive grants, and cut the budget for the program by two-thirds (it was later restored). The study also found that nonprofit organizations frequently provided compensatory education services. Congress amended the law to focus permitted activities on desegregation, and subsequent regulations prohibited use of funds for compensatory education.

The study revealed that effective organizations used citizen action strategies. Revised regulations for the program reflected this finding in requiring that such organizations have experience in working effectively with other community organizations. Thus, the study contributed to impact by identifying an important problem and by pointing to effective methods of assisting desegregation. Identification of successful practices is frequently valued, and more examples appear below.

Desegregation: Magnet Schools

These schools have special resources designed to attract students to the school and to facilitate voluntary desegregation. The study revealed that people were attracted to the schools. However, magnet schools were found to be most effective as part of an overall desegregation plan, and OE was found to be awarding funds to school districts with poor desegregation records. Generally, the program was found to be growing out of proportion to actual desegregation activities.

Congress essentially ignored these negative findings and doubled the 1980 appropriation for the program. The Appropriations Committee noted, "A recent evaluation shows that magnet schools are an effective tool in helping to improve community attitudes toward schools," (Senate Appropriations Committee, 1979, p. 107). The program was popular with Congress, perhaps because it gave the appearance of handling desegregation in a positive way. The Administration, on the other hand, asked for a recision of the 1980 appropriation, and for reduced funding in 1981.

Some might term the budget impact a misuse of the findings by the Congress. However, it should be understood in light of the debate over magnet schools' acceptability. The findings contributed to decisions and impacts because they strengthened arguments in favor of the schools. It is up to decisionmakers to draw implications from the findings, even when evaluators would argue for a different set. As Weiss (1973) notes, "politicians have a different model of rationality in mind."

Desegregation: Survey of ESAA TV Viewership

ESAA funded public television series to facilitate desegregation. The shows were aimed at minority children of school age. A survey of viewership found, however, that less than one third of the children interviewed recalled seeing even one show. The managers of the program claimed they were already aware of the viewership problem before the survey was completed, and they had already taken actions to improve viewership. Because management change was under way, the survey rapidly became obsolete. This may have happened because of poor communication and planning, bad luck, or because a survey made management sensitive to the problem and eager to solve it.

Educational Innovation: Federal Programs Supporting Educational Change

This study surveyed school districts implementing innovations supported by OE. Although some districts started innovative projects to meet needs, others did so merely to obtain federal money. These projects fell apart, as did serious innovations, if they did not have proper planning and the commitment of participants. The study also identified factors associated with the continuation of projects after federal funding ended. The House report noted these findings in amending the law for Title IV-C, which funds adoption of innovations in school districts. Projects would be funded for a maximum of five years, with a declining federal share.

This evaluation was one of the first to cast light on the problems of implementing changes in a situation of local control. The revision in our thinking about local control is probably the most profound effect of this study and has earned it much citation. The management uses of the study, at both federal and state levels, were probably most important as government officials learned they must invite, not mandate, change in schools. In fact, California passed two laws to encourage local development of needed reforms.

Educational Innovation: Project Implementation Packages

This study revealed that, even when explicit instructions are given for replicating a successful project, modifications will be necessary when project models are adopted by new school districts. According to interviewees, the study challenged the assumptions of federal managers who had believed complete replication would happen. The federal management of Title IV-C thereafter funded project developers to provide individual technical assistance to schools adopting their projects. Together with the study of Federal Programs Supporting Educational Change, this study revealed to federal funders the nature of implementation of local projects and the extent to which ideas could feasibly be transferred.

Educational Innovation: Exemplary Career Education Projects

Effective ways of teaching career education were identified in this study. The Joint Dissemination Review Panel reviewed the ten evaluated projects and decided to approve funding of seven for dissemination and adoption. Six of the projects accepted this funding, and there have been many adoptions of these innovations throughout the country. "Impact", in the final analysis, consists of these adoptions but dissemination of the project models is also impact, as we defined it.

Vocational Education: Facilities Survey

This survey, the first comprehensive information on vocational education facilities, provided a data base used by federal researchers in the areas of sex equity, youth unemployment, and education finance. Federal managers used the data to plan. It served its intended audience: managers and researchers.

Vocational Education: Sex Equity

This study concluded that students continued to be concentrated in classes stereotyped as "appropriate" for their sex. An immense amount of public interest was generated by this conclusion, but we were able to locate only one, unconfirmed use of this information at the time of our study. Moreover, the evaluation was seriously misused, in that people drew negative implications for state sex equity coordinators, when their positions had only been authorized one year before the study concluded. An additional problem was that clearance by HHS took almost two years. The study had only been available for six months when we investigated its use. It may well have been used later—such events take time for people to digest the information and put proposals for change on the agenda.

Higher Education: The Fund for Postsecondary Education

An evaluation of this fund concluded that, by many measures, it was highly successful. The evaluation was used to justify authorizing an

increased level of funding, and the fund was transferred from the Education Provision Act to the Higher Education Act, "to give the legislative visibility deemed appropriate by the committee" (House of Representatives, 1979, p. 56). We cannot say that the evaluation, strong praise though it may be, was the sole motivation for the increased budget. It nevertheless contributed by strengthening arguments in favor of an increase.

Higher Education: Campus-Based Aid

Because this survey involved four student aid programs and provided information on student characteristics, it was more useful than existing data that examined only one program at a time. The American Council on Education produced statistical analyses of the data to use in its own policy positions on student aid. The data showed a need for a change in the BEOG program. Together with other groups, ACE proposed a change in the program that "was substantially incorporated into" a House bill (House of Representatives, 1979, p. 18). At the time of this study, the bill had not yet become law. This example illustrates how an agenda, set in part by evaluation, can emerge from any of a number of sources—provided the data are made available. In this case, management information that was shared with a lobbying group supported a perceived need for change.

Higher Education: Upward Bound

This evaluation which was a follow-up of an earlier study followed the progress of poor and minority students through high school and college. It concluded that Upward Bound was successful in getting more poor and minority students to attend college. Several federal agency staff believed that the positive evaluations of this program contributed to increases in the budget. This is likely because each year Congress appropriated more than the Administration requested. Moreover, the budget, which had been level for four years, began to increase steadily after the release of the first report. However, we were unable to find a direct linkage to budget impact, beyond these beliefs.

At the time of our study, lobbying groups in postsecondary education were pressing for a change in the Upward Bound program,

and according to one respondent it was highly likely that their proposal, which used the Upward Bound data, would change the poverty criterion for the program. Also, regulations were being written that employed criteria for awards based on outcomes that evaluations had shown to be measurable.

Follow Through: Planned Variations Study

After expenditures of over $20 million in evaluation over a decade, this outcome evaluation concluded Follow Through was not superior to regular classroom instruction, and that it was questionable whether there was any clearly superior Follow Through model. We could locate at best "serious consideration" of the findings in the press, for the "back to basic" movement. When the study was in process, it also contributed to the state of the art of evaluation itself (Rivlin and Timpane, 1975; General Accounting Office, 1975; Cronbach et al., 1980). The information was probably not the kind that policy and program audiences could readily use. The program had a vocal constituency that marshalled support against any threat. For reasons other than the evaluation, the Department of Education had been seeking budget cuts. Outcome information could not easily be used in this politically impacted setting.

Follow Through: Exploratory Evaluation

This evaluation revealed conflicting views of the mission of the program that needed resolution, and disagreement about management objectives as well. The central Follow Through Office was found to have no means of producing effective services or evaluating their effectiveness. The study recommended several management changes that were implemented. For example, the Follow Through Office was reorganized, and the sponsors of project models were given a new role in local service and knowledge production. This study capitalized on people's unhappiness with the program and skirted the effectiveness issue to focus on the central management problems that were feasible to address. The existence of the program was a given; improvement of existing arrangements was the focus, and for this reason, impact was obtained.

Bilingual Education

The evaluation of this program revealed, among other things, that Hispanic students in bilingual classes were outperformed in English proficiency by Hispanic students in regular classes. In addition, less than one-third of students in bilingual classes had limited English-speaking ability and did not leave the program when they did become proficient. These findings stirred a great deal of political opposition and criticism, but Congress nevertheless used them in justifying several amendments. The law was strengthened to indicate that measures must be developed to determine when children no longer needed assistance. The definition of the target population was expanded to include those limited in reading and understanding, the measures used in the study. In addition, Congress limited the number of English-speaking children in bilingual classrooms. Finally, a five year limit was placed on federal support for any one bilingual project.

Some believe that this study contributed little to policy on bilingual education. On the contrary, one can argue that policy began to change, though not in a bald, public way. After methodological critiques and political opposition by bilingual education advocates, Congress sidestepped a redefinition of program goals. However, it took steps to make sure that English proficiency would be emphasized. In directing the bureaucracy to develop criteria for children to enter and leave the program, Congress passed on the hot potato. OE took steps to develop these criteria, but they are necessarily controversial.

National Day Care Study

This study provided persuasive evidence that increasing the number of children per staff member in a day care center would reduce cost without affecting the quality of care. Regulations for day care programs used the study's recommended age categories and adopted one of its three policy options for specifying group size and staff to child ratios. The study found that specialized training was linked to better care, though education and experience were not. Requirements for staff reflect these findings. One of four recent studies on day care, the National Day Care Study helped to resolve a 10 year debate in Congress

over the stringency of regulations, which some states could not meet. One can usually expect a finding to be used that maintains cost can be reduced without affecting quality of service.

Impact Aid

Three options were developed for the future of a program to recompense school districts for the presence of nontaxable Federal property in the community. The HEW Administration selected one set, consisting of five recommendations, as part of the Administration's legislative proposal for ESEA. Although staff said the study was discussed, Congress did not adopt the Administration's proposal. HEW had been trying to cut back on the Impact Aid program for years, because some districts receiving funds were not greatly affected by federal property. Cut-backs, however, are in obvious conflict with members' needs to keep constituents happy.

Indian Education

This study described projects funded by the Indian Education Act. Several officials mentioned that it had not been particularly helpful, merely "listing what was out there." They would have preferred a study of effectiveness, or one that identified the kinds of projects that were useful to Indian students. However, the study was cited in justifying an Administration budget proposal for the program, and possibly, a legislative proposal to increase per-pupil expenditures where low numbers of Indian children affect the success of programs.

DISCUSSION

Evaluations do contribute to policy and program decisions, and to actual changes. We were able to trace numerous contributions to law, regulation, management, and budgets. This is not to say that evaluations alone caused either the decisions or the changes. It is to say that evaluations can be seriously considered, and they can influence the content of decisions. They can sometimes even motivate a decision.

After a decision is made, findings can support arguments in its favor and thus contribute to impact. But many other considerations go into decisions, and many other forces determine program changes. In highlighting some of these contributions, we tried to indicate the forces at work in each situation.

Because our findings contradict the truism that evaluations are not used, we must reassess our methods and state why we should be believed. First, although our sample was purposive, we examined almost half of the eligible studies completed in two years by the Office of Education's evaluation unit. Even if the rest of the studies contributed nothing to decisionmaking (and we know some of them did), the extent to which OE's evaluations contributed would still be a striking disproof of current beliefs. We do not maintain our findings generalize beyond two years of OE studies. The study that samples evaluations across many policy sectors and many years remains to be done. Yet it is now reasonable to believe that such evaluations might contribute something to decisionmaking.

Second, it is always possible our respondents embroidered on the truth, and decisions were not influenced by the evaluations. This may be a serious problem, since we performed the study for Congress. However, we verified the claims of each respondent as extensively as possible, through independent statements and documentation. A conspiracy would have to stretch from Congressional staffers themselves, to independent researchers, to bureaucrats who had varying perspectives and no particular love for evaluation. We discovered only two instances in which respondents flat-out disagreed with each other, and in both cases the likely explanation is a simple mistake. In summary, our methods, though flawed, are adequate to support the contention that these studies contributed to decisions and to impact on policies and programs. We now compare our findings to the literature on knowledge use in policymaking.

HOW ARE EVALUATIONS USED?

Our findings conflict with current beliefs about the contributions of evaluation in important ways. The beliefs are:

(1) Evaluations are rarely used in decisions.

(2) When they are used for decisions they are a minor input in most cases.

(3) They are used largely as window-dressing to legitimate decisions that were already made on other grounds.

(4) When they are used in decisions, the decisions are usually unimportant.

(5) Evaluations' major use is to enlighten policymakers about the nature of policy and programs.

Extent of Use in Decisionmaking

The belief that evaluations are rarely used in decisions comes primarily from the study of Patton and his colleagues (1977). They interviewed the decisionmaker, the evaluator, and the project officer responsible for each of twenty evaluations of health programs. Fourteen of eighteen responding decisionmakers and thirteen of fourteen evaluators said their evaluation was used. However, in no instance did respondents relate specific findings to specific decisions. Evaluations were used as "additional pieces of information in the difficult puzzle of program action" (p. 145). They were used in a context of ongoing decisionmaking, rich with other information.

Like Patton, we found evaluations served as one of many inputs. And the extent of use we discovered was relatively similar—all studies except two were used in decisionmaking. However, we were able to relate specific findings to specific decisions—129 instances and of these, 68 instances in which the decisions were implemented in changing programs and policies. By any measure, contributions to decisions were the rule, not the exception.

Several reasons are evident for the differences between the Patton study and our own. First, Patton et al. were careful to draw a random sample of evaluations, while our sample was purposive. Conceivably, our findings are an overestimate of the extent to which typical evaluations contribute. As we noted, however, the contributions from OE's evaluation unit are still surprisingly frequent. A second difference in the studies involves the sample of respondents. Patton et al., interviewed "the person identified by the project officer as being either the decisionmaker for the program or the person most knowledgeable about the study's impact" (p. 143). In our study, contributions were scattered all over the federal education sector—typically, there was no single decisionmaker who knew every contribution. Moreover, "the decision maker for the program" in Patton's study may have been so far above the day-to-day administrative responsibility for the program that

he or she did not know the whole story on use. This is especially serious, since Rich (1979) found that subordinates do not always cite findings that are the basis for a decision. We interviewed middle-level bureaucrats who had themselves used findings. Had we interviewedsay, the head of the ESAA programs insteead of those who drafted program regulations, we would not have discovered evaluations's contribution to these regulation.

A final reason for the difference in findings is that Patton examined the health sector in 1975 and we examined education in 1980.[4] Evaluation had five more years to establish itself. Also, it has been a prevalent federal activity for a longer time in education than in health services delivery (excepting clinical trials). Many of evaluation's mistakes and subsequent lessons were made in education programs. In addition, many features of OE facilitated use. A stability was evident in the staffing of education offices that many federal health agencies have not achieved. The director of Title I services had been in office ten years, and the evaluation officer, twelve years at the time of our study. Most administrators, whether they liked evaluation or not, knew what to expect and to whom they should talk. Bureaucratic problems of rapid reorganization and failed communication hinder use (Leviton and Hughes, 1981), and are everywhere in federal health programs. Also, many respondents agreed that the Congressional staff who made use of evaluations in education were much more sophisticated on the subject than is usually the case when Congress deals with data (Zweig, 1979). Staff had dealt with the same policy issues through several Congressional cycles and were prepared to integrate new information into their understanding.

Importance of Evaluations as Inputs to Decisions

The Patton et al. study as well as some case studies (Bauman, 1976; Menges, 1978; Millsap, 1978) indicate that when evaluations are used, they are likely to be one of many inputs to thinking about policy and programs. One may therefore ask: how important an input is evaluation relative to other sources of data and other considerations? The studies of evaluation (Florio, Berman, and Goltz, 1979; Fox, 1977) and of social science generally (Caplan et al., 1975; Weiss and Bucuvalas, 1977; Weiss and Weiss, 1981) indicate that federal decisionmakers can value research information highly in some circumstances. But does it carry weight in decisions?

We can only speculate about the ultimate importance of evaluations as inputs. However, the decision contexts give us clues about importance, as do the opinions of respondents. When, for example, the study "Federal Programs Supporting Educational Change" offered new information about implementing change in schools, respondent opinions and citations of the findings in laws to foster such change allow us to make an inference that evaluation was important. In contrast, our respondents said that evaluations played only a minor (though real) part in some administration budget proposals. Importance varied greatly.

Quantifying importance is meaningless in these cases—qualitative understanding is everything. One must ask, "Important in relation to what, the need for program survival or for political visibility?" No. "In relation to gossip, newspaper headlines, and opinions?" Perhaps. Information from several sources is used in combination (Caplan et al. 1975; Rich, 1977), so can we really partition the variance due to evaluation?

Contribution to Decisions Versus Legitimation of Decisions

Knorr (1977) concluded that, although social science is used in making decisions, it is also used to legitimate decisions that are made on other grounds. Weiss (1978) notes that such use is entirely appropriate in gaining political support for a position. We agree and say such use contributes to impact. Yet our task here is to distinguish legitimation from use in decisions. Again, we can only infer the extent to which legitimation was involved. We have several bases for such an inference. Respondents tell us. Also, respondents and documents may indicate that data played a role before the appearance of a formal bill, regulation, or management order. Legitimation becomes much more an issue afterward, as the proposal is challenged in public debates (Bauman, 1976; Mitchell, 1980). Third, selective use of data provides an indication, as in the case of Magnet Schools. Yet even here, there is evidence that positive findings helped resolve a political debate that in turn led to a decision to change the budget. The findings were not just window-dressing.

Thirteen of the twenty-seven instances of "serious consideration" were attempts to gain support for a political position. Such attempts were often involved when evaluations contributed to decisions as well. However, the decisions also took their form from the evaluations, and in no case could we clearly label the instances involving decisions as mere

legitimation. For example, lobbyists for higher education made use of the Campus-Based Aid survey to argue for a change, and the data helped convince Congress about the merits of the case. As Pelz (1978) noted, persuasion and other uses are not neatly compartmentalized. They should all be expected to occur in political debates.

Importance of the Decision

Weiss (1981) suggests that when decisions are made on the basis of evaluations, they are likely to be less important ones that do not threaten political constituencies. Yet one cannot maintain that the decisions to which evaluations contributed here were unimportant. Clearly, no program died because of evaluation; clearly, other considerations went into budget decisions. But the changes in the law and regulation for Title I, bilingual education, and higher education were extremely important in their potential impact on students. Even though a small start at improving education of neglected and delinquent children was not earthshaking, improvement was the all-important and correct goal. Importance is not a matter of budgets and program survival alone. The reader should judge the importance of contributions in terms of incremental political change (Lindblom, 1968).

Direct Versus Indirect Use in Policymaking

Weiss (1977) concluded from past studies of knowledge use that an important function of social science is to provide enlightenment about policies. Findings would "percolate through the consciousness" of policymakers and eventually influence programs in important ways. Given the work of Caplan et al., (1977), and Weiss and Bucuvalas (1977), we expected many instances in which people would tell us that findings confirmed or revised their thinking about programs—serious consideration, but nothing more.

Perhaps our most surprising finding is the number of decision-related uses compared to those involving serious consideration only. Even in the Rich (1977) study, where use in decisions was common, conceptual use was more common. We can say that in our study, serious consideration was involved in all decisions and all impacts. Enlightenment was essential to many of these, as when "Federal Programs Supporting Educational Change" revealed to federal officials the nature of local implementation of innovations. Yet the linkage to action was

much more direct than we were led to expect. For example, in the National Day Care Study, recommendations translated directly into regulations. In the case of bilingual education, the actions of Congress were hardly direct, as they took a step toward redefinition of goals, and then left the hard work to OE. But the implications of the study were relatively straightforward. There was no evidence of a long "percolation through consciousness" in these instances. It is possible that the decisions in retrospect appear more straightforward then they were at the time. Or, that the "percolation" may have occurred in the long years of evaluation prior to these studies. The debates over Day Care and Magnet Schools had been well-articulated for some time, and questions about Title I performance were frequently very specific. In contrast, debates over the Follow Through outcome studies may take years to result in any decisions, and arguments on sex equity were yet to be marshalled at the time of our study.

Summary

We do not maintain that current beliefs about evaluation are completely false. Several examples in our work illustrate attempted legitimation, enlightenment, or unimportant contributions. However, the generalizations may well be false, and impede a careful study of contributions. Evaluations are certainly performed for many reasons (Suchman, 1967), but to the extent that they are "decision driven," we can reasonably expect contributions to decisions.

WHAT KINDS OF EVALUATIONS CONTRIBUTE?

With many examples available to us, we are in a better position to generalize about the kinds of evaluations that can be expected to contribute, and to speculate about the reasons for this. A most striking finding is the prevalence of contributions by implementation information, in all decision contexts. It can best be understood in light of the variation in local responses to federal policies. Our decentralized system has always allowed local governments great freedom in interpreting federal policies, and the extent of local control leads to problems of implementation (Danielson et al., 1977; Pressman and Wildavsky, 1973). Moreover, federal actors are largely insulated from the reality of local programs. This is more true of the powerful decisionmakers than the program managers, because information is selectively passed

upwards in bureaucracies (Downs, 1967). Nathan (1979) makes the point that federal decisionmakers simply do not know what is going on in local projects. It should not therefore be surprising that they would favor the revealing bits of information about local behavior in the face of their attempts at control.

Use of implementation findings by program management is easily understood, given the need to monitor program operation and to plan for the future. Yet contributions to legislation are even more prevalent in our sample. Why should legislators be interested? Part of the answer lies in Congress' need to make programs accountable to the public. Programs are more easily held accountable for operations than for effects. Also, decisions about implementation directly affect constituents. For example, when local districts exclude private schools from Title I services, constituents will complain.

Another reason that implementation findings contribute to legislation is that the implicit theories about program effectiveness assume correct implementation (Pressman and Wildavsky, 1973). When implementation is poor, as in the case of Title I services to neglected and delinquent children, Congress can act. It can also act when insights are obtained about how to improve effectiveness, as in the case of day care requirements. But when a program is found to be ineffective, with little information about improvements, Congress' choices are constrained to elimination, budget changes, or further study. Only the last is easy in the face of constituent pressures.

Outcome information contributed more often than we expected, given this reasoning. Outcomes were cited in two (possibly four) budget changes and an increase in authorization. However, we can most legitimately suspect "window-dressing" in the budget context, since changes might have occurred in any case. Program managers made no use of pure outcome information, except in the ten career education projects. OE had learned that once the lid is off such innovations, they will expand with or without evidence of effectiveness. Selecting effective projects through the JDRP before they gain a constituency is easier than defunding ineffective projects with a large constituency.

Where implementation is linked to outcomes and Congress has clues to program improvement, action is based in part on the findings. Similarly, the bureaucracy can write regulations based on such clues to program effectiveness. The case of nonprofit organizations in desegregation offers an example. The combination of implementation and outcome information is a powerful tool for policymakers just as it is for evaluators.

When used by Congress, federal administration information indicated management problems, many of which contributed to the variation in local practices. It was the NIE Compensatory Education Study that revealed these, and it was able to contribute because it was reported directly to Congress. Congress would not likely have heard about problems from a study supervised by HEW, nor would it have accepted assurances as readily that the program was well-run. When management problems were revealed within HEW, in the case of the Follow Through Exploratory Evaluation, the information was used in organizational development.

In the NIE Compensatory Education Study, the National Day Care Study, and the Impact Aid Study, assessment of federal administration borders on policy analysis. In each case, an assessment of current program operation (evaluation) was married to projections about the future, and to concrete policy alternatives. In these instances, policies were adopted directly from the alternatives set forth, and were developed in concert with the policymakers themselves. Inclusion of policy analysis appears to promote usefulness.

MODELS FOR REFORM?

Some of these studies may serve as models for evaluation reform if we accept that greater use in decisions is a goal of reform. Correct use is one goal, but we should keep the danger of misuse in mind, as well as strive for fairness, quality, and good planning. Fortunately, some of the highly utilized studies satisfy these other desirable goals. For example, the NIE Compensatory Education Study began with a planning period in which all stakeholders were consulted about important evaluation questions. This served the goals of equity as well as greater relevance.

These findings fuel the fires of certain reforms and not others. None of the reports were purely case studies, and are therefore irrelevant to the merits of using such methods. On the other hand, reliance on implementation findings reinforces our appreciation for the importance of understanding the program prior to evaluating outcomes. Some may argue from our data that outcome findings are not used; but this is not the case. Outcomes in isolation from implementation are to be avoided, perhaps, as they contributed largely in the special environment

of the JDRP. However, the power of combining implementation and outcome data for program improvement is evident. It reinforces Cronbach et al.'s (1980) claim that when evaluation is used, it is used in a formative mode.

Finally, our case studies should encourage evaluators by showing that their work does have potential utility. We hope these examples will enlighten debates about evaluation, and ground them in more realism about program and policy settings.

NOTES

1. A characterization of service delivery always accompanied information about recipients in these studies. Therefore we did not distinguish the two types of information. Some critics may also quarrel with our designating the adoption of innovations as outcome information. However, it was the goal of Title IV-C to get innovations adopted.

2. We could, for example, have used the proportion of individual findings from studies that contributed. However, decisionmakers use finding in clusters (Caplan et al., 1975; Rich, 1977), so that this unit may be misleading. Another alternative unit would be the formal actions of policy and programs: sections of law and regulations, management orders, memos, and reports, RFP's, etc. Thus in the example above, a single contribution would have been counted: "an administrative proposal." Since more than one change could occur in a single formal action, however, this unit loses information.

3. Our study ended in late 1980. We have not followed developments since.

4. Patton (personal communication) believes that greater use is to be expected in the 1980's because evaluation is a more familiar tool for the manager and policymaker.

REFERENCES

BAUMAN, P. (1976) "The formulation and evolution of the health maintenance organization policy, 1970-1973." Social Sci. and Medicine 10 (March/April): 129-142.

BORUCH, R. F. and D. S. CORDRAY [eds.] (1980) An Appraisal Educational Program Evaluations: Federal, State and Local Agencies. Washington, DC: U.S. Department of Education. (ED 192 446).

CAPLAN, N., A. MORRISON, and R. STAMBAUGH (1975) The Use of Social Science Knowledge in Policy Decisions at the National Level. Ann Arbor, MI: Institute for Social Research.

COHEN, D. K. and M. S. GARET (1975) "Reforming educational policy with applied social research." Harvard Educ. Rev. 45 (February): 17-41.

COOK, T. D. and W. E. POLLARD (1977) "Guidelines: How to recognize and avoid some common problems of mis-utilization of evaluation findings." Evaluation 4: 161-164.

CRONBACH, L., S. R. AMBRON, S. M. DORNBUSCH, R. D. HESS, R. C. HORNIK, D. C. PHILIPS, D. F. WALKER, and S. S. WEINER (1980) Toward Reform of Program Evaluation. San Francisco: Jossey-Bass.

DANIELSON, M. N., HERSHEY, and BANES (1977) One Nation, So Many Governments: A Ford Foundation Report. Lexington, MA: Lexington Books.

DOWNS, A. (1967) Inside Bureaucracy. Boston: Little, Brown.

DUNN, W. N., I. I. MITROFF and S. J. DEUTSCH (1981) "The obsolescence of evaluation research." Evaluation and Program Planning 4, 3/4: 207-218.

FLORIO, D.,M. M. BEHRMAN, and D. L. GOLTZ (1979) "What do policymakers think of educational evaluation?—Or do they?" Educational Evaluation and Policy Analysis 1: 61-88.

FOX. H. W. (1977) "Evaluation research and Congressional oversight: A case study of a general phenomenon." Presented at the annual meeting of the American Association for the Advancement of Science, Denver, CO, February.

GUBA, E. G. and Y. S. LINCOLN (1981) Effective Evaluation. San Francisco, CA: Jossey-Bass.

HILL, P. (1980) "Evaluating education programs for federal policy makers: Lessons from the NIE Compensatory Education Study," in J. Pincus et al., Educational Evaluation in the Public Policy Setting. Santa Monica, CA: RAND Corp. (R-2502-RC).

KNORR, K. D. (1977) "Policymakers' use of social science knowledge: Symbolic or instrumental?" in C. H. Weiss (ed.) Using Social Science Research in Public Policy Making. Lexington, MA: Lexington Books.

LEVITON, L. C. and R. F. BORUCH (1982) "Usefulness of evaluations: The NIE Compensatory Education Study." In press, J. of Policy Analysis and Management.

LEVITON, L. C. and R. F. BORUCH (1980) "Illustrative case studies," in R. F. Boruch and D. S. Cordray (eds.) An Appraisal of Educational Program Evaluations: Federal, State, and Local Levels. Washington, DC: U.S. Department of Education. (ED 192 446).

LEVITON, L. C. and E.F.X. HUGHES (1981) "A review and synthesis of research on the utilization of evaluations." Evaluation Review 5 (August): 525-548.

LINDBLOM, C. E. (1968) The Policy-Making Process. Englewood Cliffs, NJ: Prentice-Hall.

MENGES, C. C. (1978) Knowledge and Action: The Use of Social Science Evaluation in Decisions on Equal Educational Opportunity, 1970-1973. Washington, DC: National Institute of Education.

MILLSAP, M. A. (1978) "The uses of evaluation in federal regulation writing: A case study." Presented at Evaluation Research Society Convention, Washington, DC.

MITCHELL, D. E. (1980) "Social science impact on legislative decisionmaking: Process and substance." Presented at the American Education Research Association Annual Convention, Boston, April.

NATHAN, R. P. (1979) "Federal grants-in-aid: How are they working in 1978? Presented for Conference on "The city in stress." Sponsored by the Center for Urban Policy Research, Rutgers Univ. March 9.

PATTON, M. Q., P. S. GRIMES, K. M. GUTHRIE, N. J. BRENNAN, B. D. FRENCH, and D. A. BLYTH (1977) "In search of impact: An analysis of the utilization of federal health evaluation research," in C. H. Weiss (ed.) Using Social Research in Public Policy Making. Lexington, MA: Lexington Books.

PELZ, D. C. (1978) "Some expanded perspectives on use of social science in public policy," in J. M. Yinger and S. J. Cutler (eds.) Major Social Issues: A Multidisciplinary View. New York: Macmillan.

PRESSMAN, J. L. and A. WILDAVSKY (1973) Implementation. Berkeley, CA: Univ. of California Press.

RICH, R. F. (1979) "Problem-solving and evaluation research: Unemployment Insurance policy," in R. F. Rich (ed.) Translating Evaluation into Policy. Beverly Hills, CA: Sage.

RICH, R. F. (1977) "Uses of social science information by federal bureaucrats: Knowledge for action versus knowledge for understanding," in C. H. Weiss (ed.) Using Social Research in Public Policy Making. Lexington, MA: Lexington Books.

RIVLIN, A. M. and P. M. TIMPANE [eds.] (1975) Planned Variation in Education. Washington, DC: Brookings Institution.

SAXE, L. and D. KORETZ (1982) "Editors' notes." Making Evaluation Research Useful to Congress. New Directions for Program Evaluation 1982 (June): 1-8.

SUCHMAN, E. A. (1967) Evaluative Research. New York: Russell Sage.

U.S. General Accounting Office. (1975) Follow Through: Lessons Learned from Its Evaluation and the Need to Improve Its Administration. (MWD-75-34). Washington, DC: U.S. General Accounting Office, October 7.

U.S. House of Representatives, Committee on Education and Labor (1979) Education Amendments of 1980 (96th Congress, 1st Session, Report No. 96-520). Washington, DC: U.S. Government Printing Office.

U.S. House of Representatives, Committee on Education and Labor (1978) Report: The Education Amendments of 1978, H.R. 15. (95th Congress, Second Session, Report No. 95-1137). Washington, DC: U.S. Government Printing Office, May 11.

U.S. Office of Education, Office of Evaluation and Dissemination (1980) Annual Evaluation Report, Fiscal Year 1979.

U.S. Office of Education, Office of Evaluation and Dissemination (1979) Annual Evaluation Report, Fiscal Year 1978.

U.S. Senate, Committee on Appropriations (1979) Report on the Departments of Labor and Health, Education and Welfare, and Related Agencies Appropriation Bill, 1980. Washington, DC: U.S. Government Printing Office.

WEISS, C. H. (1981) "Measuring the use of evaluation," in J. A. Ciarlo (ed.) Utilizing Evaluation: Concepts and Measurement Techniques. Beverly Hills, CA: Sage.

WEISS, C. H. (1978) "Improving the linkage between social research and public policy," in L. E. Lynn (ed.) Knowledge and Policy: The Uncertain Connection. Washington, DC: National Research Council.

WEISS, C. H. (1977) "Research for policy's sake: The enlightenment function of social research." Policy Analysis 3 (Fall): 531-545.

WEISS, C. H. (1973) "Where politics and evaluation meet." Evaluation 1, 3: 37-45.

WEISS, C. H. and M. J. BUCUVALAS (1977) "The challenge of social research to decision-making," in C. H. Weiss (ed.) Using Social Science Research in Public Policy Making. Lexington, MA: Lexington Books.

WEISS, J. A. and C. H. WEISS (1981) "Social scientists and decision makers look at the usefulness of mental health research." American Psychologist 36 (August): 837-847.

ZWEIG, F. M. (1979) "The evaluation worldview of Congressional staff," in F. M. Zweig (ed.) Evaluation in Legislation. Beverly Hills, CA: Sage.

Laura C. Leviton is Assistant Professor and Policy Analyst at the Health Policy Institute, Graduate School of Public Health, University of Pittsburgh. She is currently a W.K. Kellogg Foundation Fellow and is co-author with Thomas Cook and William Shadish of "Evaluation" in the forthcoming Handbook of Social Psychology *(3rd Edition). Her interests include health services research, health policy, and the relation between evaluation and policy analysis.*

Robert F. Boruch is Director of the Program in Methodology and Evaluation Research, Department of Psychology, Northwestern University. He is the co-editor with R. F. Riecken of Experimental Testing of Public Policy, *and with others, of* Social Experimentation. *His scholarly papers concern evaluation policy and the development of methods for planning and evaluating social programs.*