CSE
MONOGRAPH
SERIES
IN
EVALUATION

10

EVALUATION IN SCHOOL DISTRICTS:

ORGANIZATIONAL PERSPECTIVES

Edited by
Adrianne Bank
Richard C. Williams
with the assistance of
James Burry

Center for the Study of Evaluation
University of California • Los Angeles

EVALUATION IN SCHOOL DISTRICTS: ORGANIZATIONAL PERSPECTIVES

CSE MONOGRAPH SERIES
IN EVALUATION

SERIES EDITOR

Eva L. Baker

Center for the Study of Evaluation
UCLA Graduate School of Education
University of California, Los Angeles
Los Angeles, California 90024

EVALUATION IN SCHOOL DISTRICTS: ORGANIZATIONAL PERSPECTIVES

Edited by
Adrianne Bank
Richard C. Williams
with the assistance of
James Burry

Center for the Study of Evaluation
UCLA Graduate School of Education
University of California, Los Angeles, 1981

CSE MONOGRAPH SERIES IN EVALUATION

NUMBER

1. Domain-Referenced Curriculum Evaluation: A Technical Handbook and a Case Study from the MINNEMAST Project
   Wells Hively, Graham Maxwell, George Rabehl, Donald Sension, and Stephen Lundin
2. National Priorities for Elementary Education
   Ralph Hoepfner, Paul A. Bradley, and William J. Doherty
3. Problems in Criterion-Referenced Measurement
   Chester W. Harris, Marvin C. Alkin, and W. James Popham (Editors)
4. Evaluation and Decision Making: The Title VII Experience
   Marvin C. Alkin, Jacqueline Kosecoff, Carol Fitz-Gibbon, and Richard Seligman
5. Evaluation Study of the California State Preschool Program
   Ralph Hoepfner and Arlene Fink
6. Achievement Test Items—Methods of Study
   Chester W. Harris, Andrea Pastorok Pearlman, and Rand R. Wilcox
7. The Logic of Evaluative Argument
   Ernest R. House
8. Toward A Methodology of Naturalistic Inquiry in Educational Evaluation
   Egon G. Guba
9. Values, Inquiry, and Education
   Hendrik D. Gideonse, Robert Koff, and Joseph J. Schwab (Editors)
10. Evaluation in School Districts: Organizational Perspectives
    Adrianne Bank and Richard C. Williams with James Burry (Editors)

This Project was supported in whole or in part by the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This monograph is embedded within a larger scope of work funded by the National Institute of Education and carried out over a six-year period by the Center for the Study of Evaluation. This work began with the identification of large and small school districts who had organized their evaluation or testing services into a centralized unit. Then, CSE proceeded with the development, distribution, and analysis of a national survey describing the work, staffing, and budgeting of such units. Several series of case studies followed the survey. A guidebook for district managers who wish to improve instruction through the use of testing and evaluation results will be forthcoming from CSE's most recent work.

As editors of this volume, we have been involved in these efforts from their inception. But, as in any extended endeavor, many people have contributed both directly to this volume and indirectly as influences on our thinking.

Catherine Dillon Lyon was the project director of the national survey. She, working with James Burry, Linda Cullian, Lynn Doscher, Pam McGranahan, Janet Sutton, and Clinton B. Walker, designed the instrument and then collected and wrote up the data.

Organized in parallel with that survey effort was a panel coordinated by David O'Shea, and consisting of the four individuals who authored the chapters in this monograph—Oscar Grusky, Charles O'Reilly, David O'Shea, and Lynne Zucker. They made suggestions for the survey design and conducted additional analyses after the initial survey findings were published. Egon Guba, Professor of Education at Indiana University, played an important facilitating role in initiating the panel's activities while he was a Visiting Scholar at CSE.

At CSE, Eva L. Baker, the Center's Director, provided the administrative support and energy essential for the overall project and the work of the faculty panel.

Aurelia Arroya and Katherine Fry patiently and accurately typed and retyped the many drafts.

Mary Ann Millsap, our NIE Project Monitor, supplied continuing encouragement and invaluable editorial guidance on the early drafts of the papers.

Finally, and especially, we want to acknowledge the cooperation of the evaluation unit directors who completed the questionnaires and shared their working knowledge with us.

To all those who participated in one way or another over the years in bringing this aspect of CSE's work to fruition—we extend our sincere thanks.

Adrianne Bank
Richard C. Williams

PREFACE

Mary Ann Millsap

National Institute of Education

Over the past decade, increased emphasis and reliance have been placed on the educational roles of testing, assessment, and evaluation. Issues raised in these domains reflect a broad range of concerns, from methodological concerns about the validity and reliability of measures to political concerns about the appropriate and equitable use of technical information, from assignment of students in classrooms to federal policy in the appropriations of billions of dollars for educational programs. They thus have relevance for an equally broad range of decision-making levels—local, state, and federal.

For the past five years, the National Institute of Education (NIE), the research and development arm of the U.S. Department of Education, has funded a number of studies on these issues. Included among its research activities are studies on how testing and evaluation information is used in the schools, on the impact of testing programs, on the relationship of testing and evaluation to instruction, and on new methods for assessing educational outcomes. Workshops and conferences on testing and evaluation for educators and the general public have been conducted, resulting in publications targeted for lay audiences as well as for the research community.

The research agenda arose in response to the unprecedented growth in testing and evaluation in schools over the past 15 years and because of the absence of systematic information about the nature and impact of those activities. The Center for the Study of Evaluation at UCLA recognized early the need to explore how school districts use testing and evaluation information. With NIE funding, CSE undertook in 1977 the first national survey of evaluation and testing offices in school districts. This work was primarily descriptive in nature, focusing on how such offices were organized, what their budgets were, what activities they undertook, and whom they identified as their major clients. Working in conjunction with faculty from the University of California's Los Angeles and Berkeley campuses, CSE incorporated into the survey

several items relating to school districts as social organizations and the roles of staff within such organizations. These items, buttressed with the considerable sociological literature on organizations, form the basis of the papers contained within this monograph.

The monograph makes an important contribution to the field of testing and evaluation use. The papers provide considerable insight, from a theoretical, sociological perspective, into the incentives and disincentives for using testing and evaluation information in schools and into the role conflicts faced by the providers of such information. From a careful reading of these papers, the editors extrapolate the changes and modifications that are needed so that technical information can be more effectively used within school and school district contexts.

The monograph also complements other NIE supported research in testing and evaluation. Both the Center for the Study of Evaluation and the Huron Institute are conducting intensive case studies in school districts around the country, particularly in those school districts identified as successfully tying evaluation and testing information to administrative decisionmaking and instructional reform. The rich data bases from these studies are providing much needed information about the issues faced by a variety of educators—superintendents, school boards, principals, program managers, and teachers—and about how testing and evaluation information has assisted in addressing these issues. The findings from this research are being translated into technical assistance strategies to assist other school districts in the better use of evaluation and testing information.

In addition to funding these studies, NIE also supports an annual grants program in testing and evaluation, designed primarily for school districts to conduct their own research on testing and evaluation issues within their school districts.

By funding interdisciplinary work such as this monograph on the organizational factors influencing evaluation and test use, by increasing field based research on how districts use evaluation and testing information, and by encouraging school districts to submit proposals of their own design into NIE-funded grants announcements, NIE is attempting to increase the utility of testing and evaluation information for more effective schools.

Introduction

This monograph has grown out of the Center for the Study of Evaluation's (CSE) on-going interest in the conduct of educational evaluation, how it operates in school districts, and how it contributes to educational practice. The title—*Evaluation in School Districts: Organizational Perspectives*—we believe, conveys the orientation of the monograph and its content.

The monograph focuses on evaluation and how it occurs within a specific setting: school districts' central administrative offices. This narrows our interest from evaluation-in-general to evaluation-in-operation within this special context. Thus, this volume examines the relationships between evaluation activities and school district organizational features that impinge upon such activities. Research on evaluation methodology and evaluation use often pays lip service to the need to understand the context within which evaluation occurs. But there are very few research studies which attempt to relate organizational variables such as decision making, internal administrative structures, role definitions, and the nature of school districts' "technical core" to the way in which district personnel conceive, conduct, and utilize educational evaluation. Throughout, the monograph offers multiple perspectives on district organizations rather than advocating a single perspective. The chapters represent a range of sociological views on evaluation and its function in relation to other school district operations. Each of the chapter authors worked with CSE staff members on conceptualizing a national survey of school district evaluation units and in analyzing the resultant data. Findings from the survey itself, which describes evaluation work as it is carried out in school districts, are summarized later in this introduction.

While the chapters reflect this common data base, they differ in approach. Each of the authors agreed either to use the CSE survey data as a stimulus for reflecting on evaluation utilization or performed additional analyses of the data to examine their own hypotheses about evaluation utilization and the functioning of educational evaluation units. The authors are Charles O'Reilly III,

UC Berkeley Graduate School of Business Administration; Lynne Zucker, UCLA Department of Sociology; Oscar Grusky, UCLA Department of Sociology; and David O'Shea, UCLA Graduate School of Education.

O'Reilly's comprehensive review of the decision making research literature reminds evaluators of the preeminence of the human factor in decision making. School district decision makers, like those in other organizations, are not rational, consistent, and logical; they are not disposed to make decisions on the basis of "hard" data. Instead they usually consider many factors, only one of which may be evaluative data, and they often arrive at decisions that reflect organizational predilections for their own self-interests. Given the decision makers' organizational constraints, they commonly and understandably use evaluations, or selected portions, to justify what they already had planned to do, to interpret data to be supportive of achieved outcomes. Sometimes they even refute or ignore data that run counter to their preferences.

Zucker, through her reanalysis of the CSE survey data, illustrates how school districts utilize evaluations to perform a "signaling" function to the external environment. This signaling behavior, typical of organizations which must respond to societal demands, leads to evaluations that are primarily reactive to the information needs of external funding agencies. Because of the potential or actual conflict among these external dimensions, she asserts, most evaluations appear ambiguous and offer only a narrow range of information; the usefulness of such evaluations for informing districts' internal decisions about educational programs is thereby diminished.

Grusky combines his analysis of CSE's survey results with the literature on role conflict and role ambiguity to point out characteristics of complex organizations that evaluation unit directors must consider if their work is to influence organizational decision making. His analysis contains organizational, structural, and behavioral implications for individuals who lead school district evaluation units and for others who may be involved in designing and conducting local evaluations.

O'Shea, finally, discusses some of the antecedents of current evaluation practice and places the CSE survey findings in a historical context. He reminds us of the somewhat arbitrary manner in which legislative requirements for evaluation took form. O'Shea further develops a theme explicit or implicit in each of the preceding chapters: that school districts are institutional rather than

technical organizations and thus their evaluation imperatives are externally focused. This point, when coupled with the fact that causal relationships between classroom instruction and learning outcomes are erroneously presumed by current evaluation practice further constrains the utilization of evaluation findings. Given this contextual background, O'Shea argues that it is reasonable to expect that current evaluation practice will provide information having only limited impact on the way in which school districts make internal instructional decisions. He concludes with some recommendations for overcoming this problem.

This monograph, and the CSE work out of which it grew, is unique. The project brought together a number of sociologists who first contributed to the design of empirical survey work being done on school district evaluation practice and next reflected upon the findings from that work from the standpoint of theories and literatures not common to educational evaluators.

The survey was initiated because CSE recognized that educational evaluation, as it occurred within school districts, was virtually undocumented by researchers. Although many school districts were themselves conducting, or employing others to conduct, evaluations in response to federal and state reporting requirements, very little summary information was available about how they were managing their evaluation tasks. CSE was aware that some of the larger, and a few smaller, districts had centralized their evaluation functions into research and evaluation units. Both CSE and the National Institute of Education were curious about how these evaluation units were staffed, funded, and organized. Consequently, CSE was contracted by NIE to investigate these issues at the national level.

The survey was intended, primarily, to describe the centralized evaluation capacity being developed within districts. But it seemed to us that the organizational characteristics of school district central offices, as well as the organizational characteristics of evaluation units themselves, merited additional scholarly attention from specialists who could describe the surrounding context and weigh its influence on evaluation practice and usefulness. Consequently, we organized a team of individuals to represent both viewpoints.

The early meetings of this team were devoted to broad discussions of research strategy and to specific items that should be included in the questionnaire seeking information about school district evaluation units. The bulk of the responsibility and day-to-day work involved in designing and administering the question-

naire and initially analyzing its results was performed under the supervision of Catherine Dillon Lyon, who at that time was project director. During the formulation of the survey the contributions of the team were primarily supplementary and supportive. Later, after CSE project staff had prepared a descriptive study from the survey responses (Lyon, et al., 1978; Williams, 1979), four members of the team—the four chapter contributors in this monograph—undertook their own independent analyses of the data. What was unique about our joint CSE-specialist efforts, then, was their interdisciplinary nature.

In the remainder of this introduction we briefly review some recent organizational perspectives for understanding school district functioning, as well as recent evaluation perspectives relating to issues of utilization. Then we describe the CSE survey of school district evaluation units and summarize its major findings.

In the body of the text, we precede each chapter with a short abstract summarizing the major arguments and their significance. The volume concludes with a synthesis and brief discussion of the implications of this work.

Recent Perspectives on School Districts as Organizations

The understanding of organizations and of administrative decision making has undergone important changes in recent decades. Earlier, the dominant, idealized view of organizations was that of a rational, clearly designed bureaucratic structure staffed by workers whose main preoccupation was the organization's successful functioning. Other organizational manifestations, such as informal structures, staff self-interest, decision maker preferences, favoritism, and the like were considered deviations from the ideal and summarily described as system "noise." The job of the good manager was to reduce that "noise" and keep the organization running smoothly, without irrational interference. Thus, managers were taught that they would be successful if their organizations were properly organized and managed—using such concepts as span of control, specified goals, and clearly written rules and regulations. Writings in the educational administration literature clearly reflected this dominant organizational viewpoint.

Over the years, however, other perspectives on organization have been gaining acceptance. What was previously viewed as irrational, e.g., the dominance of administrator preference in decision making, or goal displacement, or informal subsystems

became increasingly viewed as natural outcomes of organized human endeavors. Organizational theorists have sought to embody such concepts and activities within their thinking and formulations. Another way of describing this phenomenon is to consider the field as moving from a relatively simplistic certainty to a relatively complex uncertainty. But as of now, there appears to be no all-encompassing organizational theory that adequately captures the complexity of organizational functioning.

In this monograph the authors refer to a wide range of notions and concepts about organizations, and the reader unacquainted with the organizational literature may not be familiar with them. Among the more important of those concepts are "loose coupling" (Weick, 1976), "institutional" vs. "technical" organizations (Meyer & Rowan, 1977), and "organized anarchies" (March & Olson, 1976).

Loose coupling refers to the degree to which various parts of an organization (its individual staff, divisions, units) are linked to one another. If management acts to establish a new employee procedure, for instance, will it be implemented at the employee level? If not, that part of the organization can be considered to be loosely coupled. If the desired effect is quite immediate, it is tightly coupled. For example, a superintendent might require that all teachers use certain materials in their classrooms. If, after a reasonable period of time, his or her preferences were not routinely implemented, the district's functioning as regards that specific order would be considered loosely coupled. It should be noted that organizations can be simultaneously loosely coupled in some matters and tightly coupled in others.

Institutional and *technical* organizations differ with regard to whether they primarily articulate societal norms or a specific technology. Technical organizations, such as manufacturing enterprises, implement a specific technology. Since management has full knowledge of all processes required to transform raw material inputs into finished product outputs, the organizational structure can be tightly coupled. If a technical organization has an evaluation unit, that unit can relate deficiencies in quality of output back to the specific production processes responsible.

In contrast to technical organizations, institutional organizations implement societal expectations in areas such as religion, welfare, and education. These provide service in areas that the community wants addressed, but there is a limited body of tested knowledge upon which the managers of those organizations can draw to

implement procedures guaranteed to produce desired objectives. Implementation, therefore, commonly is delegated to certified professional practitioners whose activities are only loosely coupled to management. Neither managers nor evaluators usually attend directly to the central technical problems but rather deal with them indirectly or symbolically. School district administrators, for example, may define a good school as fulfilling social expectations regarding credentialled teachers, requisite numbers of library books, and a published instructional continuum. How the credentialled teachers actually teach the instructional continuum is dealt with only incidentally. An evaluation unit may prefer to devote its resources to demonstrating compliance to external expectations rather than to assessing the effectiveness of instructional procedures relative to pupil achievement outcomes.

Organized anarchies is a term used to describe phenomena actually occurring in many organizations. These phenomena are quite counter to what rational bureaucratic theory would lead one to expect to find. The term attempts to portray something that may be organized but may not be rational according to standard organizational theory. Proponents of this view (for example, March & Olson, 1976) argue that many organizations, either occasionally or usually, exhibit this kind of behavior to a lesser or greater degree. The behavior can be described as "choices looking for problems," "solutions looking for issues," and "decision makers looking for work."

These concepts—loose coupling, institutional organization, and organized anarchies—together suggest quite a different view of organizations than is pictured by rational bureaucratic models. But these three concepts do not describe organizational aberrations; rather they represent the organizational world as it really is.

The extent to which school districts operate as loosely coupled, institutional, and anarchic organizations has profound implications for thinking about and perhaps re-considering the role and value of educational evaluations in districts' decision making. At the very least, it probably means that evaluators have to come to understand the organizational constraints within which educational administrators make decisions. This understanding is crucial if evaluation information is intended to be used in the districts' decision-making process.

Recent Perspectives on Evaluation

Within the past five years, as evaluation activities have been increasingly mandated at policy levels and conducted at local levels, evaluation scholars have been concerned about utilization. Some scholars examining evaluation utilization view the phenomenon or its absence as a special case of the study of knowledge utilization (Caplan, 1975; Weiss, 1979; Boruch, 1980). They are concerned with how the production of research knowledge in specialized fields, including education, gets transmitted in a useable form to those who must make policy decisions. The audiences of principal concern to these researchers are primarily, though not exclusively, federal and state legislators and administrators who must allocate scarce resources among competing program alternatives.

There are also individuals working on evaluation utilization from the point of view of defining the problems inherent in the practice of evaluation and improving its procedures and practices (Alkin, Daillak, & White, 1979; Patton, 1978). These researchers view evaluation as a service to decision makers, and reason that if such services are not perceived as useful to clients—at whatever level the clients may be—then the evaluation profession needs to reexamine the way in which it goes about its work.

A third group (Apling, Kennedy, & Newman, 1980; Thompson & King, 1981; Williams & Bank, 1981) approaches evaluation and the data it generates as one of many sources of information residing within a dynamic and often politically volatile organizational context. They examine evaluation as one among many ongoing processes and they consider the use or non-use of evaluation findings in local districts in terms of relationships among those processes.

It is only within the past five years that this last group, which is interested in examining the organizational (as distinct from the knowledge production or the professional) aspects of evaluation utilization, has become a major force within the evaluation research community. We believe that CSE's work, underwritten by NIE and expressed in this monograph, is an important contribution to this development. The monograph depicts the factors affecting organizational decision making and describes some of the tensions existing among actors within an organization. Since educational decision making usually does not follow the evaluator's rational model of problem statement, alternative solution generation and weighting, followed by the selection of an optimum solution, the

monograph provides some tentative advice for both evaluators and decision makers which may help them recognize each others' constraints and mutually work towards increasing the usefulness of evaluation information.

The CSE National Survey

The national survey was undertaken to broaden our understanding of how evaluation activities contribute to the renewal of public education at the school district level. Its primary function was descriptive. As an indication of the lack of basic information about school district evaluation functions, it took CSE staff more than six months simply to locate the appropriate respondents' names and addresses and generate a basic mailing list.

Survey respondents were finally selected through a two-stage process. First, letters were sent in the fall of 1978 to all 750 superintendents in school districts with 10,000 or more students and to a 50 percent sample of the 573 school districts with 5,000–9,999 students. All of the larger districts and 81 percent of the smaller ones responded indicating whether or not their district had an evaluation unit. Next, in spring, 1979, a questionnaire was sent to the 320 large school districts (10,000 or more students) and to the 74 smaller ones identified as having evaluation units. A total of 263 unit heads (or 64 percent) returned the survey questionnaire.

The findings include the following:[1]

> Only 43 percent of the districts enrolling 10,000+ have a central office responsible for program evaluation. The existence of an evaluation office is partially a function of district size. Almost all (89 percent) of the metropolitan districts, 59 percent of the large, and 33 percent of the medium districts have an evaluation unit.
>
> Formal school district evaluation is a relatively young and growing field in education. Only 15 percent of the existing evaluation offices were established before 1965. Since 1970 the creation of these offices has accelerated, with 51 percent newly organized or reorganized from other units (testing, research) since then.
>
> Generally, evaluation offices are more likely to be in one of the typical lines of authority, rather than directly reporting to the

[1] From Lyon and Doscher (1979). See also Lyon, Doscher, McGranahan & Willians (1978) and Williams (1979).

Superintendent. Thirty-seven percent of the respondents report directly to the Superintendent, 44 percent report through an intermediary, and about 20 percent through two or more intermediaries. Thirty-nine percent are located in instruction offices, leaving 61 percent located elsewhere (administration and support services, for example).

Central evaluation offices are primarily though not likely to be exclusively in charge of formal assessment. Only 26 percent report exclusive responsibility for evaluating both state/federal and local programs.

Evaluation office activities are dominated by school district personnel rather than consultants. Almost half of the units do not spend any money on consultants. Two-thirds of the respondents reported that their personnel resources are not adequate.

Most evaluation offices administer testing programs. Seventy-one percent administer district-wide norm-referenced testing (NRT) programs. Forty-eight percent administer district-wide criterion-referenced testing (CRT) programs. Districts administering both CRT and NRT programs use them for the same purposes. Approximately 75 percent agree that testing is their major method of data collection.

The average evaluation office spends approximately 40 percent of its time on instructional clients compared with 60 percent with administrators. Generally, administrative rather than instructional personnel are perceived to be using evaluation unit data. Evaluation offices typically do not take part in other administrative functions (personnel, collective bargaining, budget). Offices express little agreement on what constitutes basic evaluation practices or on the priority of various evaluation activities.

Conclusion

The preceding pages only touch upon the thematic concerns of the monograph. Each of the topics we have raised is amplified in one or more of the following chapters. Together, the authors generate theoretical propositions about the role of evaluation in educational organizations, test some of these propositions against CSE's empirical survey data, and offer valuable insights to both decision maker and evaluator. While we do not claim to have illuminated a direct path to the heart of the labyrinth, perhaps we have suggested the dimension and complexity of the maze and indicated some of the promising directions to take.

References

Alkin, M. C., Daillak, R., White, P. *Using evaluations: Does evaluation make a difference?* Beverly Hills, California: Sage, 1979.

Apling, R., Kennedy, M. M., & Newman, W. F. *The role of evaluation and test information in public schools.* Cambridge, Mass.: The Huron Institute, 1980.

Boruch, R. F., & Cordray, D. S. (Eds.). *An appraisal of educational evaluations: Federal, state and local agencies.* Northwestern University, 1980.

Caplan, N., et al. *The use of social science knowledge in policy decisions at the informal level.* Ann Arbor: Institute for Social Research, University of Michigan, 1975.

Joint Committee on Standards for Educational Evaluation. *Standards for evaluations of educational programs, projects, and materials.* New York: McGraw-Hill, 1981.

Lyon, C. D., & Doscher, L. Executive summary, evaluation and school districts, Los Angeles: Center for the Study of Evaluation, University of California, 1979.

Lyon, C. D., Doscher, L., McGranahan, P., & Williams, R. C. Evaluation and school districts. Los Angeles: Center for the Study of Evaluation, University of California, 1978.

March, J. G., & Olson, J. *Ambiguity and choice in organizations.* Bergen, Norway: Universitetsforlaget, 1976.

Meyer, J. W., & Rowan, B. Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology,* 1977, *83,* 340–363.

Patton, M. Q. *Utilization-focused evaluation.* Beverly Hills, California: Sage, 1978.

Thompson, B., & King, J. A. Evaluation utilization: A literature review and research agenda. Paper presented at annual meeting of the American Educational Research Association, Los Angeles, 1981.

Weick, K. E. Educational organizations as loosely coupled systems. *Administrative Science Quarterly,* 1976, *21,* 1–19.

Weiss, J. A., & Weiss, C. H. What makes social science useful for public policy? Paper presented at the annual convention of the American Psychological Association, New York, September, 1979.

Williams, R. C. Studies in evaluation and decision making: Evaluation offices in small school districts. CSE Report to the National Institute of Education. Los Angeles: Center for the Study of Evaluation, University of California, November 1979.

Williams, R. C., & Bank, A. Uses of data to improve instruction in local school districts: Problems and possibilities. In C. Aslanian (Ed.), *Improving educational evaluation methods.* Beverly Hills, California: Sage, 1981.

Notes on O'Reilly's "Utilization of Evaluation Research"

Introduction

In the 1960's, the early years of educational evaluation, much attention was directed, quite properly, toward evaluation's technical aspects. The field was dominated by such questions as: What is an appropriate design to assess the effectiveness of an educational program? What statistical methods are most useful to analyze complex multi-level data? How do you write clear, measurable objectives that reflect the intention of the program developer? What are effective data collection techniques? Some of these emphases can be traced back to the origins of the evaluation movement; some can be traced to the interests of the psychologists and measurement specialists who provided the movement's early leadership.

Considerable progress has been made toward resolving these methodological concerns, although there remains much healthy debate in the field over what methods are most appropriate under what circumstances.

More recently, some evaluators have begun shifting their attention to a different set of questions which have to do with the utilization of evaluation data. The kinds of questions being asked are: Who will ever read this evaluation report? Are these evaluations to be used by decision makers and, if so, under what circumstances and for what purposes? What factors inhibit or increase evaluation use? Research on evaluation utilization has provided us with mixed answers: Evaluation reports are being used in some instances and not in others; often their use is as one source of data that fits into the decision makers' thinking; in some instances evaluations can be decisive; in many, they have more limited influence.

Those interested in trying to better understand evaluation utilization and thereby provide guidance to those seeking to improve the level of utilization, have worked in different, but complementary directions. Some view evaluation utilization as a subset of

knowledge utilization and so examine the roles of information production and transmission. Others have sought answers through further improving the technical and methodological side of evaluation. They reason that decision makers' non-utilization rests in part on their perception that evaluations are not technically sound or are poorly written or poorly timed. Others have begun to look more closely at the decision-making context and the characteristics of decision makers and the organizations they work in. They ask such questions as: What are the many things decision makers consider when reaching a decision? What personal or organizational or environmental constraints influence decision-making behavior?

The study of decision making is, of course, not a new field of inquiry. For decades, indeed centuries, scholars and practitioners alike have been fascinated with decision making and decision makers. That Herbert Simon won the Nobel Prize for his work in this field attests to its importance. Because of its obvious relevance to administrators and managers, much of the work has originated in and been encouraged by professional schools of administration and management.

Argument

Charles O'Reilly III, a sociologist in the Graduate School of Business Administration, University of California, Berkeley, is a student of organizations and decision making. In his chapter he attempts the difficult task of selectively reviewing several relevant literatures—sociology, social-psychology, decision making, administrative theory, organizational theory. He then draws out propositions that will provide insights to those interested in the evaluation utilization problem.

The main thesis of his argument is that organizational leaders and decision makers are not neutral entities who are waiting for, or even desirous of receiving, carefully constructed evaluation studies that objectively report and interpret data that will help them reach informed decisions. Instead, organizational leaders and groups in organizations have personal, political, social, and economic influences and constraints that largely determine the decisions they make. Evaluation may play a minor, and occasionally a major role in decision making, especially if the findings are consistent with what the decision makers had already concluded was the best course of action.

O'Reilly reminds us that decision making in complex organizations, such as school districts, is not usually the result of one single event nor the product of one individual. Rather, it evolves from complex and ongoing bargaining and negotiation, both formal and informal. Information, as represented by evaluation, is sometimes ammunition in a continuing battle between those factions that support or oppose a given decision. Seen in that light, it is easier to understand why evaluations can be valued or discredited, quoted or ignored, regardless of their intrinsic technical characteristics.

Contribution and Importance

To readers not familiar with this line of inquiry, approaching O'Reilly's chapter may seem a formidable task. The language, the approach, and the format may be unfamiliar. But his work makes several important contributions to the evaluation utilization question. He introduces the reader to the major research in several related fields; he spells out the complexity of the utilization problem and thereby reduces the likelihood that the evaluation field will reach simplistic solutions to complex problems; and finally, he presents a series of tentative propositions that evaluation scholars and practitioners alike should consider when conducting evaluations or inquiring further into evaluation utilization.

Evaluation Information and Decision Making in Organizations: Some Constraints on the Utilization of Evaluation Research[1]

Charles O'Reilly III
University of California, Berkeley

INTRODUCTION

The CSE project studying evaluation activities in school districts addresses the question of the extent to which school district evaluation offices are contributing to the renewal of public education. Part of this question can be answered empirically by examining data describing how local evaluation agencies (LEA's) are organized and staffed and what functions they serve (Lyon, Doscher, McGranahan, & Williams, 1978). Still another part of the question can be addressed through organizational analyses using survey data to test propositions concerning the impact of structure or conflict (see Grusky and Zucker elsewhere in this monograph). As David (1978) has documented, it appears that Title I evaluation results have not served, primarily, as a means of judging program effectiveness or as a guide to program improvement. It is also apparent that evaluation information is rarely seen as important in the context of decisions about program changes. David goes on to point out that "literature on evaluation has only recently included attempts to understand the role and use of evaluation results . . . in the realm of decision making" (1978, p. 8). Her study provides evidence of the almost total lack of utilization of evaluation data by decision makers in school districts.

Since the CSE survey data do not directly address the empirical testing of the propositions generated here, this chapter may best be seen as an attempt to make the reader aware of some factors

[1]I would like to thank John C. Anderson, Susan Resnick, and Bill Glick for helpful comments on an earlier version of this manuscript.

which may preclude evaluation information as a force for change. From this perspective, local evaluators may be able to anticipate what aspects of their evaluation will be seen as convincing or useful by various decision-making audiences.

BACKGROUND

In the past several years, evaluation researchers and practitioners have become concerned about the apparent non-utilization of evaluation information. For instance, Davis and Salasin (1975) acknowledge that even methodologically sound research is often not used. Because many people accept the proposition that the purpose of evaluation research is to produce information primarily for use by decision makers, their concern about decision makers' non-utilization of evaluation findings calls into question the very purpose of evaluation.

There is no true consensus on whether this perceived lack of utilization is accurate. There is evidence supporting and evidence attacking the claim of non-utilization. For example, a number of authors have noted that the Head Start program continues, in spite of the results of ten or so years of research demonstrating that it has had almost no effect on cognitive abilities (e.g., Goldstein, Marcus, & Rausch, 1978). Similar instances of what appears to be non-utilization can also be found. Ward and Kassebaum (1972), for example, report that program administrators in a state correctional system's counseling program systematically ignored evaluation data showing that the program was ineffectual with respect to the specified outcomes, decided to extend the program and, in certain cases, to make it mandatory.

A number of reasons have been offered in explanation of such non-utilization:

- A lack of rigor in the design and conduct of the evaluation study (e.g., Goldstein et al., 1978).
- So much rigor that the research fails to assess the true purpose of the program (e.g., Cox, 1977).
- Reporting outcomes which are not helpful (such as negative findings), non-timely reporting of results, or providing reports which are too long and filled with jargon (Alkin, 1975; Cox, 1977; Goldstein et al., 1978).
- Communication problems between evaluators and program administrators (Cox, 1977; Goldstein et al., 1978).

Arrayed against these arguments and examples are counter-examples and claims that utilization does, in fact occur. In a survey of federal evaluators, Patton (1978) reported that 78 percent of those surveyed felt that evaluation studies had an impact on their programs. He, and others, argue that a broader definition of "utilization" must be used in order properly to assess the impact of evaluation research. When the definition of utilization is extended to include impacts such as providing additional credibility for a program, verifying managers' suspicions, and generally providing additional information, then, it is argued, the utilization of evaluation information is more widespread than is commonly believed (Levine & Levine, 1977). The problem, according to this view, is primarily that the definition of "utilization" is too narrow and "overly rational" (Patton, 1978, p. 40).

Whether utilization is or is not problematic depends on the definition of the term, on what is meant by "rational," and on the evidence the discussant uses in arriving at a conclusion; on what information the discussant seeks out or is exposed to, how closely he or she chooses to attend to the various arguments, and whether he or she has a vested interest in the outcome of the dispute. From this perspective, the utilization or non-utilization of evaluation information depends on information availability and use as well as the organizational processes and pressures which encourage or discourage the production, transmission, and use of information by relevant decision makers.

This chapter analyzes the utilization of evaluation research from an organizational behavioral perspective. To accomplish this, I will present a simplified information-processing model of decision making. With this orientation, the utilization or non-utilization of evaluation information may be seen as a function of two primary factors: (1) the organizational context in which decision makers function, including an awareness of operating goals and incentives, and (2) the information processing constraints on the decision makers, including availability of information, costs of search, and the manner in which information is perceived and processed. Drawing on the research literature related to information use and organizational decision making, a series of propositions will be generated that suggest conditions under which evaluation information is likely to be used, or not used, by decision makers.

AN INFORMATION PROCESSING MODEL FOR DECISION MAKING

Decision making is, simply put, the act of choosing among alternatives. Ideally, the decision maker is presumed to have a set of values or evaluative criteria, the perception of a problem which requires action, a number of potential alternative solutions, and a calculus for comparing alternatives and estimating the likelihood of attaining certain outcomes given certain alternatives. With *perfect* rationality, the assumptions leading to choice generally include complete information about the alternatives, knowledge of the probabilities associated with different alternative-outcome links, a consistent ordering of preference among outcomes, and a selection mechanism which maximizes the value attained by a choice. However, with *bounded* rationality, the limits on both the information and cognitive processing abilities of the decision maker are recognized; that is, the decision maker is intendedly rational but acting with limited computational abilities. In both cases, however, the choice process is one of information assimilation and use, suggesting that an information processing model may be a fruitful way to examine the decision-making process.

Consider the simplified model of decision making presented in Figure 1. In this schema, the decision maker faces a problem or situation requiring a choice, generates potential alternative solutions, assesses the probabilities that a given alternative will lead to certain outcomes, and develops a preference ordering among outcomes. In this view, information and the ability to process it are paramount. Information processing is required in all phases, i.e., to define the problem, develop alternatives, estimate probabilities, and order outcomes.

The emphasis in this model is *not* that it is a complete or literal representation of the decision-making process. A variety of similar models have been proposed (e.g., Cohen, March, & Olsen, 1972; Janis & Mann, 1977; MacCrimmon, 1974), and there is evidence suggesting that such models may be oversimplifications (Witte, 1972). Mintzberg, Raisinghami, and Theoret (1976), for example, traced 25 complex organizational decisions and characterize the process as a plurality of sub-decisions without a simple sequential relationship such as indicated in Figure 1. Nevertheless, the model in Figure 1 is useful in that it emphasizes the centrality of information and information processing in the decision process. It also suggests a framework for organizing and investigating the limitations and constraints on organizational decision making.

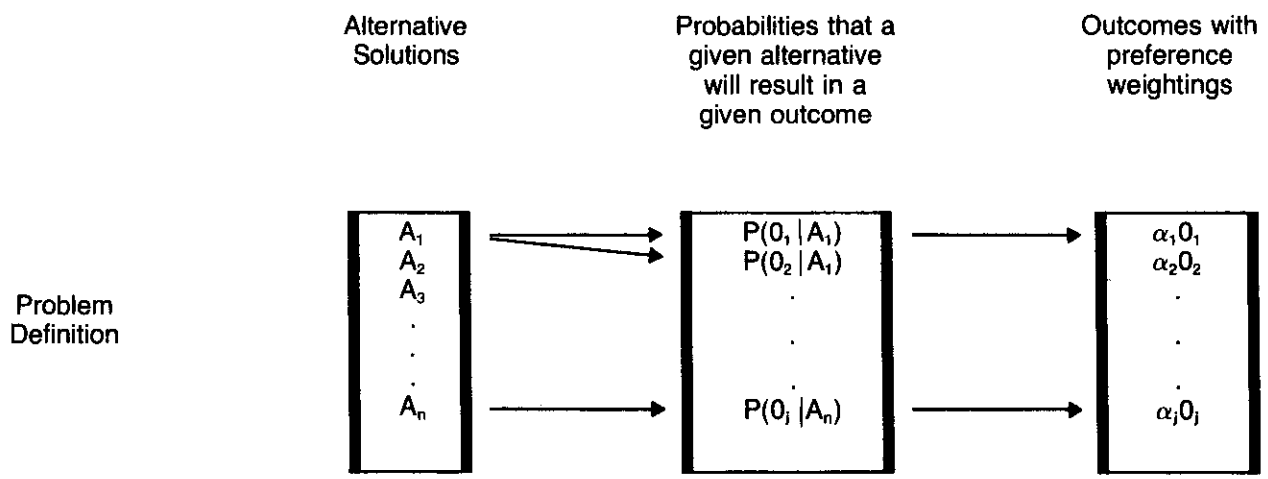

Figure 1

Simplified Model of the Decision-Making Process

In a typical formulation, decision makers encounter or are presented with problems which require choices, e.g., continue funding for a project or reallocate funds to other purposes, or allow the hiring of additional staff. The decision maker then considers a variety of alternatives for solving the problem. Each alternative is examined and some subjective weight or probability estimated that selection of given alternatives will lead to particular outcomes. Typically, the decision maker is not indifferent among the set of outcomes but has preferences reflected by weights for at least some of the outcomes. Given that information processing limits exist, it may be that the decision maker will not have complete knowledge of the alternatives, probabilities, or outcomes. In general, however, the decision-making process is presumed to operate from left to right. Under the traditional notions of rationality ascribed to by evaluation researchers, it is assumed that the decision maker will search for unbiased information about the various components in the model, and that the weights attached to various outcomes are determined by organizational goals; that is, the decision maker will attempt to make the choice such that the maximum net benefit accrues to the organization or the agency.

But is this left-to-right progression an accurate description of how decisions are made? Several authors have proposed that, in organizations, problems are seldom clearly defined and alternatives become known only after certain outcomes are preferred (Cohen et al., 1972; Weick, 1977). It may be, in fact, that the decision-making process in organizations begins with the preference ordering for outcomes as a set of rather fixed constraints. Simon (1964) suggested this when he proposed that organizational goals be viewed as constraint sets. Certainly, in organizations individuals and subunits typically have vested interests. The loss of resources, status, and power is to be avoided. Even the uncertainty caused by change may be resisted. On the other hand, increases in resources, for example, more funding or more staff, are typically preferred. Decision makers are indifferent about outcomes only when they are unaffected. In fact, decision makers may become participants in a decision-making process to insure that they *remain* unaffected.

That organizational decision makers have strong preferences for certain outcomes is well documented (e.g., Pfeffer & Salancik, 1977, 1978). These preferences, discussed at greater length later in this chapter, reflect not only organizational and subunit goals, but also individual desires and concerns such as promotion and

salary. Because these preferences exist, it follows that managers or administrators are unlikely to remain passive when decisions affecting them are to be made. Conflicts may develop when decisions can result in gains and losses to the participants. Under these circumstances, alternatives which lead to undesired outcomes may become unacceptable to certain participants. In addition, as Thompson (1967) observed, uncertainty threatens rationality. Managers, especially those in power, attempt to eliminate potentially disruptive uncertainty. Information, in this context, becomes a political resource. Data which support desired outcomes are sought out, while information which supports opposite views is to be rebutted by questioning its accuracy, for example, by obtaining other information supporting a countervailing view, or by impugning the credibility of the source of the threatening information. Information, then, is a commodity used for a variety of purposes. Under some circumstances it may be used as a basis for decision making; in others, as corroboration for decisions already made, and in still others, for symbolic reasons. Information, in this milieu, is not a fixed substance, but one which may be selectively perceived and processed.

Thus, the argument being made is twofold. First, in organizations, participants typically have preferences for outcomes which reflect organizational as well as individual goals. These outcomes act to define a set of constraints which result in the decision process moving from right to left with reference to Figure 1. Second, because of the potential for disagreements among participants, the entire decision process may be one of bargaining and negotiation as various actors attempt to pursue their interests. In this political process, information becomes a potentially important, or threatening, commodity. It is generally not perceived of as "objective." Rather, depending on the nature and importance of the goals sought, decision makers may systematically search for supporting information while ignoring other types.

It is this organizational context in which evaluation research and evaluation units, whether at local, state, or federal level should be examined. Both the evaluation unit and the information it produces may be assets or liabilities for decision makers.

A further distinction reflecting categories of evaluation information is necessary for the purposes of this chapter. For example, one can distinguish between formative and summative evaluation, with the former referring to information focused on improving the workings of a given program and the latter more concerned with

its overall impact or effectiveness. Formative evaluations are typically used to provide information to administrators to improve the conduct of the program and, as such, may be less threatening than a summative evaluation which may dictate the continuation or discontinuation of the entire project. We can make still another distinction between information collected for routine compliance with funding agency guidelines and non-routine information used to assess goal attainment. The former information is often critical in that continued funding is predicated on demonstrated compliance and substantial efforts and resources are often required to collect such information. This information, however critical, is not typically used to assess program performance. For the purposes of this chapter, "evaluation information" or "evaluation research" will refer primarily to the more non-routine, summative aspects of evaluation research. While some of the material developed may be relevant to an understanding of routine or formative evaluation, the primary focus here is on information about non-routine or summative events.

The remainder of the chapter elaborates these views and develops a series of propositions concerning circumstances under which evaluation units and information are likely to be effective. First, the concept of "rationality" as it applies to organizational decision making and information use is discussed. Second, the impact of organizational goals and incentives on the model present in Figure 1 are considered. Finally, drawing upon a diverse body of research on cognitive information processing and decision making, a number of potential biases in information processing which may affect the use of evaluation information are explored.

RATIONALITY AND THE DECISION-MAKING PROCESS

As indicated previously, the notion of rationality implied by Figure 1 is not the comprehensive rationality of economic theory in which unbiased decision makers use perfect information to maximize utility according to some completely specified and ordered preference set. Instead, decision makers begin with preferences and select actions based on imperfect expectations about their effect upon future preferences. These actions, further, may include limited and focused information search and selective perception and processing of information. Thus, "rationality," as used here, does not even refer solely to the notions of bounded rationality in which choices are made by decision makers who use imperfect

information in a satisfying manner to maximize goal attainment (March & Simon, 1958). Instead, the appropriate calculus for Figure 1 appears to be one which March (1978, p. 592) refers to as "contextual rationality" in which choice behavior is embedded in a complex of other claims on the attention of decision makers and other structures of social and cognitive relations. Organizational decision makers, in this view, are pursuing multiple objectives subject to a variety of pressures and constraints, and often with considerable ambiguity surrounding the choice process. Under these circumstances, preferences for outcomes may be the least ambiguous component of the decision process, more certain than the definition of the problem, the range of feasible alternatives, or the probabilities associated with various alternatives. In this situation, it is argued that decision makers are likely to take action which both reduces their uncertainty and helps them achieve desired outcomes (e.g., search for supportive information or selectively interpret signals as favorable to a preferred outcome). Lindblom (1959) offers some support for this in his observation that the selection of goals and the empirical analysis of the needed actions to obtain the goals are not distinct from one another but are closely intertwined. Since ends typically come before means and may be known with greater certainty, it is likely that the search for appropriate means will be highly focused.

Since individual decision makers are known to be limited in their ability to solve problems, organizational routines are often established to increase the likelihood that individuals will behave in a traditionally rational way. This focus is on the effectiveness of *procedures* used to make choices, what Simon refers to as "procedural rationality" (1978, p. 8). These routines are established, in light of the limitations suggested by bounded and contextual rationality, to emphasize rational search procedures and, insofar as possible, to ensure that decision makers have complete information. This procedural rationality may easily include provisions for information and control systems, carefully prescribed review processes, and even mandated program evaluations. The explicit attempt is to ensure that the organizational context promotes an approximation to comprehensive rationality by decision makers. These manifestations of procedural rationality may be of considerable symbolic importance (e.g., Meyer & Rowan, 1977), but of limited efficacy if decision makers are only contextually rational.

Such a reformulation of traditional notions of rationality is hardly new. For a number of years economists, political scientists, and others have acknowledged that "perfect rationality" is not an apt

description of real-world decision making (e.g., Etzioni, 1967; Floden & Weiner, 1978; Gershuny, 1978; Lindblom, 1959). What is, perhaps, relevant here is that some of the disillusionment with social science evaluation can be partly attributed to at least an implicit acceptance of the assumptions underlying traditional rationality. It is worth re-emphasizing that in an organizational context, unlike what is implied in the traditional rational model, goals are often ill-specified or lack consensus, information may be incomplete, ambiguous, and imperfect, and decision makers may be pursuing multiple competing objectives and lacking the time and computational abilities necessary to adequately utilize the available information. It may even be, as suggested here, that the decision itself is driven by the solution rather than the problem. Acknowledging these constraints, let us now turn our attention to: (1) some of the relevant organizational properties alluded to by the notion of contextual rationality, and (2) the cognitive processing limitations suggested by the notion of bounded rationality.

ORGANIZATIONAL IMPACTS ON DECISION MAKING

Organizational Power and Evaluation

To understand the impact of contextual influences on evaluation research, let us first consider the evaluation unit in an organizational context; that is, consider the evaluation unit as one group among many subunits. As Pfeffer and Salancik (1978) have noted, subunits within organizations may be viewed as actors competing for resources in a political arena. Power, in this milieu, becomes a critical determinant of the unit's ability to acquire scarce and critical resources (Pfeffer & Salancik, 1974).

Power, the crucial ingredient in this contest, may be envisioned as resulting from the subunit's ability to control critical contingencies; that is, power stems from the ability to reduce the primary uncertainty which faces the organization. As formulated by Hickson, et al. (1971), power or the ability to control contingencies and reduce uncertainty may vary according to: (1) how critical the uncertainty is which the subunit can reduce, that is, how central the contingencies are for the overall functioning of the organization, (2) how effective the subunit is in reducing the uncertainty, and (3) how "substitutable" the function served by the subunit is; that is, how easily the uncertainty reduction can be obtained from other sources within the organization. Subunits that are successful

according to these criteria are likely to be powerful. Those units that do not act to reduce important uncertainty for the organization, or whose ability to reduce uncertainty is not uniquely held, are unlikely to exert substantial influence within the organization.

Subunit power has important ramifications for an understanding of the utilization of evaluation information for several reasons. First, Pfeffer and Salancik (1974) have shown that power is directly related to a subunit's ability to obtain scarce and critical resources, for example, budget allocations and staff. Without these, it may be difficult for an evaluation unit to perform effective evaluations. Second, Pfeffer and Salancik (1977) have also demonstrated that power is used to define the criteria used in decision making. More powerful subunits may successfully emphasize those criteria on which they compare favorably. This ability to define or specify evaluative criteria may have direct impacts on program evaluation, allowing powerful programs either to specify the basis on which they are to be evaluated and thereby insure their continued success, or to invalidate unfavorable findings by redefining *a posteriori* the evaluative criteria. Without power, members of the evaluation unit are likely to find that they lack discretion in selecting the criteria and rejecting others' definitions (e.g., Meltsner, 1976), or that they adhere to the more universalistic standards of research only to have their product labelled as irrelevant or ignored by decision makers.

A final ramification of power which has been noted is the tendency of those in charge to institutionalize their power and to resist changes which might alter their position. Goldstein, et al. (1978) describe how groups often desire evaluation research to satisfy external demands to "do something," but simultaneously are looking for the results to justify established policies and procedures. Decision makers, it seems, are more receptive to research conclusions that fit nicely into established policies. This tendency—and ability—of powerful units to maintain their position acts to slow down the process of change and may easily contribute to both a desire and an ability on the part of powerful subunits to attend to selectively (or ignore) utilization information, depending on their interests.

These findings suggest several propositions with respect to the influence of evaluation units and the utilization of evaluation research:

Proposition 1: Results of evaluation research are more likely to be used by decision makers when:

a. The evaluation unit is powerful compared to the unit being evaluated.
b. the evaluation information is not available from other sources; i.e., the control of uncertainty is not substitutable.
c. the evaluation study is on a program requiring a substantial proportion of the agency's resources; i.e., the project evaluated is central to the agency's functioning.
d. the criteria used for evaluation are criteria which the unit being evaluated has accepted, i.e., there is agreement between the evaluating unit and the one being evaluated on what is to be evaluated.

Organizational Goals, Incentives, and Evaluation

If, as suggested in Figure 1, the use of information may be a function of the outcomes preferred by a decision maker, it is important to consider how these preferences might develop. In an unconstrained setting, predicting the preferences or tastes of decision makers is problematic. With the exception of some global notions of hedonic preferences or the idea that, in general, people will choose to be "better off" rather than "less well off," it is virtually impossible to predict *a priori* what set of outcomes a given decision maker is likely to value. In organizations, however, the situation is far more predictable. Decision makers exist in settings which act to constrain both the range of outcomes which might be preferred in situations where there is a choice and the preferences for particular outcomes within this reduced set. These contextual constraints are typically imposed through systems of goals, control systems, and incentives. Decision makers are usually constrained first by the nature of the job; that is, organizational or subunit goals are imposed along with the responsibility to work toward attaining these goals. To ensure that this is accomplished, some form of control system is typically used which allows superiors in the hierarchy to monitor the achievement of subordinate decision makers. Upon discovery of a variance, sanctions can be applied. Thus, the organizational context acts to constrain decision makers to pursue a limited number of goals and to reward or punish them for success or failure.

Without digressing into a lengthy discussion of goals and goal formulation processes (cf., Mohr, 1973; Perrow, 1970), it should be noted that this process involves the translation of often vague,

non-operational, long-term goals into more specific, short-term outcomes. This activity may, as has been empirically demonstrated (e.g., Pfeffer & Salancik, 1978), be accomplished through a political process of negotiation, bargaining, co-opting, coalition formation, and the garnering and application of power. The process may also result in the adoption of symbolic and non-operative goals whose purpose is often to satisfy external audiences and to provide for minimum constraints on organizational actors (e.g., Perrow, 1970; Simon, 1964). Meyer and Rowan's (1977) interpretation of the symbolic nature of the operation of schools is an illustration of these goals. The concern in many public organizations with accountability is another goal, perhaps more relevant as a symbolic effort to reassure taxpayers, than as an operative goal relevant to effective and efficient organizational functioning.

With respect to evaluation, Orlans (1973) has noted that enabling legislation is often written in vague terms in order to gain sufficient backing from legislators to insure passage. However, since program goals are not well-defined, it is difficult to develop measurable criteria, despite the fact that the legislation often mandates an evaluation component. The lack of well-specified goals may encourage the political bargaining over criteria described by Pfeffer and Salancik (1977). It may also result in evaluations which opponents claim are not appropriate. Program Head Start legislation, for example, offered no delineation of goal priority or concrete objectives. Consequently, results of evaluation studies which showed no program impact were often ignored by decision makers as inappropriate or assessing only a part of the program's intent (Gordon & Morse, 1975).

The result of the translation of organizational goals into short-term operative subunit goals, while perhaps an imprecise and ambiguous process, has several important ramifications for an understanding of the use of information by organizational decision makers. First, even when the goals are somewhat ambiguous, objectives act to focus the attention and efforts of those responsible. Second, although the goal may be imprecisely defined, control systems are almost always developed which include measures of objective outcomes. These observable outcome measures act as highly salient foci against which performance may then be assessed. A large and persuasive body of research has documented the motivating effect of simple goal setting (e.g., Latham & Yukl, 1975). For example, studies of budgeting (e.g., Wildavsky, 1974), public employees (Blau, 1964), students (Baum & Youngblood,

1975), homeowners (Erez, 1977), have all demonstrated that the mere setting of goals can act to direct behavior. Other research has also demonstrated that once an individual becomes committed to an outcome through public actions, there is a tendency to continue to pursue the same goal, even if preceding efforts met with failure (e.g., Salancik, 1977; Staw & Ross, 1978). These findings suggest that in an organizational context, decision makers will probably develop rather explicit preferences for outcomes.

Finally, and most important, these preferences are solidified through the operation of organizational incentive systems: decision makers are rewarded for pursuing certain ends and punished for others. The pervasive impact of incentive systems should not be underestimated. Kerr (1975) provides a number of instances in which employees behaved in seemingly contradictory ways, often appearing to do the opposite of espoused goals. For example, directors of orphanages were found to establish policies that worked against the placement of children in foster homes, and acted to keep them in the orphanage; universities routinely establish incentive systems that militate against encouraging high quality teaching; sports teams often reward individual performance when a team-oriented effort is required. Upon analysis, the results were seen not to be contradictory but to be entirely consistent with the incentive systems; that is, people in organizations typically do those things for which they are rewarded. In Kerr's (1975) example, for instance, orphanage directors who succeeded in placing children in foster homes would be "rewarded" by cuts in staff, resources, and prestige among peers. In a direct test of the hypothesis that decision makers would respond to the operative control system, Harrell (1977) demonstrated that subordinate decision makers would follow their superior's lead and make decisions using similar criteria. This effect persisted even when superiors began making decisions which were contrary to the official policy.

Obviously, the goals of survival and obtaining resources are probably more important to program administrators and staff than to independent evaluators. Information which suggests that the program is not effective or efficient is likely to result in negative sanctions against those running the program. Clearly, there may be conflicts between the program administrator's goal of survival and the evaluator's goal of an accurate assessment. The incentives for those with vested interests in the program may be to devise strategies to minimize the collection, dissemination, and use of such unfavorable information, even if such actions go against some "objective" assessment of overall worth.

These findings are applicable in understanding how decision makers develop preferences for outcomes: subunit goals are made salient, measurable, and rewarded, control systems provide feedback and sanctions which act to focus attention on achieving certain ends, failure to attain the desired outcomes may result in the loss of important organizational rewards such as promotion, pay, and status, as well as actual punishments such as demotion or termination. As individuals in organizations become more committed and less able to leave the organization, the importance of the outcomes is increased. Since there is seldom complete consensus among individuals and subunits on the goals or outcomes to be sought, the possibility of conflict and competition arises. Under these circumstances, it is possible that one person's gain is another's loss; hence, it is unlikely that decision makers will easily give up preferences for certain outcomes if it means personal loss.

Some indirect evidence of these effects may be seen in a review of 93 evaluation studies by Gordon & Morse (1975). These authors found that evaluation researchers with some affiliation with the project being evaluated were far more likely to report the project as successful than were non-affiliated evaluators (52 percent to 14 percent). Similarly, there was a greater tendency for non-affiliated researchers to report program failure than was the case among affiliated evaluators (32 percent to 14 percent). While not suggesting directly that affiliated evaluators deliberately biased their studies, the results are consistent with the hypothesis that incentives may lead to conscious or unconscious biasing of information.

The set of acceptable or preferred outcomes on which a decision maker is likely to focus is, in part, made salient through the specification of a control system which assesses attainment of the assigned goals, incentive systems which sanction certain actions, and the committing effect of previous behaviors. These factors suggest the following proposition:

> *Proposition 2:* Results of evaluation research are more likely to be used by decision makers when:
>
> a. the evaluation information can be used to assess achievement of quantifiable goals.
> b. the evaluation information is fed into a well-articulated and operating control system which includes an effective set of incentives; i.e., measured performance can be sanctioned.
> c. the evaluation information does not recommend actions which are incompatible with the existing control system's ability to monitor and sanction.

d. the evaluation information results from studies performed by those who have a vested interest in the continuation of the project.

INFORMATIONAL IMPACTS ON DECISION MAKING

As suggested by Figure 1, information plays a central role in the decision-making process. Before information can have an impact, however, it must reach, and be processed by, the relevant decision makers. In an evaluation context this means that the evaluation report must not only be available to users but must also be read, understood, believed, and acted upon. Failure by decision makers either to obtain the information or to process it can result in non-utilization. This suggests the need to examine two separate steps in the dissemination process: (1) how decision makers acquire information, and (2) how material is processed cognitively. A substantial body of research is available which suggests the limitations and biases of organizational decison makers with respect to information acquisition and use.

Information Acquisition

Accessibility of Information. A large number of laboratory studies of information and decision making have documented the intuitively reasonable conclusion that better quality information is generally associated with improved decision-making performance (e.g., Porat & Haas, 1969; Streufert, 1973). Unfortunately, several authors have also noted that the majority of these studies may be overly structured when compared to the real-world situations they are meant to model (e.g., Connolly, 1977; Winkler & Murphy, 1973). Thus, while we have support for the importance of information in decision making, we need to examine the process by which decision makers acquire information before concluding that the mere availability of better quality information insures improved performance.

When we examine the evidence, doubts about the information-decision making link emerge. Decision makers of many types appear to be noticeably biased in their procurement of information. For example, in a direct test of the impact of accessibility and quality of information on information source use, O'Reilly (1979) found that although decision makers recognized information sources of high quality, they used sources which provided lower quality

information but were more accessible. O'Reilly (1979) explains these results in terms of the costs involved in obtaining information from less accessible sources. Given that the decision makers were under time constraints and subject to numerous interruptions, it may have been that they were simply unable to seek out higher quality information when it came from less accessible sources. Similar findings have been reported about the information-seeking behavior of physicians (Menzel & Katz, 1955), scientists (Gerstberger & Allen, 1968), policymakers (Clausen, 1973), and managers (Mintzberg, 1973).

This bias towards accessible information is also reflected in managers' strong preferences for oral as opposed to written information (e.g., Dewhirst, 1971) and for information from trustworthy or credible sources (e.g., Beach, Mitchell, Deaton, & Prothero, 1978; Giffin, 1967). Research in these areas has shown that managers typically prefer shorter oral reports to longer written ones. Interestingly, there is also evidence that when obtaining information in this manner, managers may judge the validity of the information, not on the facts of the matter, but on the credibility of the source. This may lead to the acceptance of a piece of information as "true" or "false" depending on how much the recipient trusts the sender. Clausen (1973), for instance, noted that Congressmen frequently cast votes on legislation based not on an understanding of the deeper issues but on the advice of a trusted colleague. This behavior is not necessarily bad, but simply reflects the inability of a Congressman to be fully informed on all issues.

The research on source credibility also suggests that it may be that it is the "safeness" or trustworthiness of the source, more than expertise, which determines whether information is believed (O'Reilly & Roberts, 1976). The ramifications of this finding for understanding the utilization of evaluation research may be important. Evaluation units, by their nature, may be perceived of as not sharing the same objectives as other subunits; i.e., not being trustworthy in terms of source credibility. These units may be seen as objective rather than sympathetic; that is, expert but not trustworthy. Thus, there is some possibility that the information produced may be used less by decision makers than information from "safer" sources. David (1978), for example, found that the feedback of evaluation data which included personal explanations by evaluation staff was much more likely to be utilized than a report that was merely delivered.

Together, these biases may lead decision makers toward preferences for information from particular channels which they may characterize as accessible and trustworthy, and which provide condensed treatments of complex issues. Information from these sources, as will be suggested later, may be concrete and easily assimilated, but not necessarily detailed or of the highest quality. With respect to the utilization of evaluation research, this suggests the following proposition:

> *Proposition 3:* Results of evaluation research are more likely to be used by decision makers if they are:
> a. readily accessible to final decision makers
> b. summarized
> c. presented orally
> d. from a source deemed as credible i.e., trustworthy.

Information Sign. An additional bias noted in studies of information acquisition by decision makers is a tendency to avoid information which may suggest undesirable consequences. Janis and Mann (1977) offer several examples of this behavior describing how politicians and policy makers, when faced with unpleasant alternatives, will avoid exploring ominous implications of desired courses of action. Several examples of this sort are available in studies showing that program administrators have exhibited tendencies to dismiss negative findings (e.g., Borgatta, 1966; Carter, 1971). Goldstein et al. (1978, p. 33) report similar instances in which evaluation results were selectively interpreted as offering support for a politically favored position. This willingness to avoid acquiring pejorative information and to seek out or selectively process favorable items has direct ramifications for evaluation research. As suggested earlier, decision makers often have vested interests in certain outcomes. This may predispose them to seek out information which supports their position. David (1978) quotes a decision maker as saying, "I want information to justify the expansion of the program" (p. 17). Bear and Hodun (1975), for example, found that subjects were likely to recall items of information that were confirmatory to their position rather than those that were contradictory. Interestingly, this bias extended to recalling some contradictory items as supportive and even to recalling missing data as confirmatory for a preferred position. In an interesting laboratory study, Morlock (1967) demonstrated that it required significantly *less* information for subjects to arrive at a decision favorable to their position than to arrive at a decision considered

to be against their interests. In an organizational setting, groups of like-minded decision makers may exaggerate these biases toward selective perception and actually act collectively to censor or derogate information in opposition to their desired ends. Janis (1972), labelling this process as "groupthink," provides a number of retrospective accounts in which groups acted to bolster desired opinions and exclude contrary ones. For instance, decision making by President Kennedy's advisory committee during the Bay of Pigs incident was characterized afterwards by the suppression of doubts, creation of feelings of unanimity and invulnerability, and an unwillingness to risk conflict within the group. Johnson (1974), in a study of group decision making, provides a nice example of this tendency to seek unanimity and avoid conflict. She hypothesized that executives would make a less desirable, but acceptable, short-run decision to avoid generating conflict with others in the group. Using 49 businessmen across ten situations she discovered that although subjects could identify the ideal decision for each situation, their overwhelming tendency was to make a sub-optimal decision in order to avoid conflict. David (1978) again quotes a user of evaluation data as saying, "I look at test scores mainly to confirm my own impression. If they differ, my impression counts" (p. 16).

These biases, that is, tendencies to avoid information which suggests undesirable consequences, to seek out supportive information, to require less supportive information to arrive at a favored decision, and a desire to avoid making decisions which will generate conflict, suggest the following proposition:

> *Proposition 4:* Results of evaluation research are more likely to be used by decision makers if the information:
>
> a. is supportive of the outcomes favored by the decision makers.
> b. does not lead to conflict among the set of relevant actors.

Proposition 4 appears most appropriate when, as Janis and Mann (1977) suggest, the group is cohesive and members value their associations. Under these circumstances, it may be that group members will be more apt to suppress conflict in the name of group unanimity. When evaluation information is available to competing factions or decision makers belonging to differing groups, the following proposition is suggested:

> *Proposition 5:* Results of evaluation research are more likely to be sought out and used by decision makers under competing or conflicting conditions when:

a. Information supportive of an undesired position is available to actors seeking different outcomes.
b. The decision-making process is drawn out over a long period of time which allows alternative positions to be presented.

Communication of Information. A final set of potential biases on information acquisition by decision makers in organizational settings is related to the communication of information in organizations. Several authors have noted that communications in organizations may be withheld or distorted (e.g., O'Reilly & Pondy, 1979; Wilensky, 1967). Subordinates have been found to be biased toward passing superiors information which reflects favorably on the subordinate and suppressing unfavorable information. O'Reilly (1977) discovered that subordinates who did not trust their superior were willing to suppress unfavorable information even if they knew that such information was useful to their boss. Other investigations have shown similar results. For instance, Pettigrew (1972) documented how a single individual, acting as a gatekeeper for information flowing to a policymaking group, was able to determine the outcome of a purchasing decision by carefully allowing only certain types of information through to the decision makers. Plott and Levine (1978) demonstrated how, through the arrangement of the agenda of a meeting, outcomes could be determined in advance. Lowe and Shaw (1968) provided evidence that departments systematically inflate and bias budget requests to support claims for increased resources. In a slightly different vein, Kaufman (1973) showed how subordinates learned not to pass certain items of information upward in the hierarchy because superiors, upon learning of these, would be required to act in ways contrary to the subordinate's self-interest. A classic example of this system-induced distortion is provided by McCleary (1977) in a study of how parole officers report clients' violations, observing that parole officers under-reported deviant behavior to their supervisors. As noted by Kaufman (1973), subordinates are often punished for providing accurate reports. In McCleary's study, this resulted in incidents being reported only when the information sent upward would enhance the subordinate's career.

Numerous other examples are available which attest to the fact that information is often selectively filtered and distorted as it is communicated in organizations (e.g., Allison, 1971; Janowitz & Delany, 1957, etc.). Such filtering and distortion appears to result, most often, from individuals or groups attempting either to gain

desired outcomes (such as increased resources or power) or to avoid their loss. When considered in conjunction with the biases toward reliance by decision makers on short, oral reports from trusted sources, the impact of distorted information may be heightened.

Clearly, these biases have important consequences for the transmission and use of evaluation research since interpretation of the results of evaluation studies may be biased either to support or refute a particular position. If, as March and Simon (1958, p. 165) note, "inferences are drawn from a body of evidence and the inferences, instead of the evidence itself, are then communicated," the opportunity for subtle distortion is magnified. Findings which have undesirable consequences may be withheld by superior gatekeepers also. Direct evidence of this effect is recounted by Coleman (1972) who describes how HEW attempted to minimize the impact of an EEO study because its findings were inimical to the interests of some other HEW agencies. Here we find decisions being made to suppress or alter information both to minimize conflict and to avoid undesirable outcomes.

It should be noted that not all distortion in organizational communication is necessarily intentional. With any transmission from one individual to another there is almost always some degradation of the message (e.g., see Campbell, 1958). Receivers of information recall certain parts of messages and forget or minimize others. This unintentional distortion, due to differences in cognitive tuning, may be increased when the communication occurs between groups who use different vocabularies, are sensitive to different goals and constituencies, or are using different criteria for determining what is important. Janis (1972), for example, describes how information during the attack on Pearl Harbor was not transmitted because senders, unaware of the broader picture, did not perceive certain information as important. Evaluation researchers, whose concerns are often somewhat different than administrators,' may not fully appreciate the concerns of users of their studies. The use of a different vocabulary and the requirements for communication through several hierarchical levels may lead to evaluation reports which administrators see as not timely, too technical, focused on issues which are not central, and generally unconvincing (e.g., Alkin, 1975; Cox, 1977). Such opinions on the part of administrators, coupled with previously mentioned biases such as reliance on accessible and credible information sources, make it clear that evaluation information may not have the impact expected by evaluation researchers.

Acknowledging that information may be blocked or altered during communication within organizations suggests the following proposition:

Proposition 6: Results of evaluation research are less likely to be useful when:

a. transmitted through several intermediate links in a communication network rather than delivered directly to the relevant decision makers.
b. transmitters of the information are likely to suffer personal or organizational losses from the message.
c. senders and receivers do not trust each other.
d. the information is transmitted to decision makers in another agency or organization.
e. the information is transmitted to decision makers in a different functional group within the same organization.

Information Processing

Discussion to this point has focused on biases which may affect the acquisition of information by decision makers in an organizational context. Once the information has been acquired, however, it may still be that individuals will fail to process it accurately. A diverse and well-developed body of research is available which documents these limitations in cognitive information processing (e.g., Slovic, Fischhoff, & Lichtenstein, 1977). Two general themes from this research are relevant to an understanding of the use of evaluation research by decision makers: (1) factors which limit one's ability to assimilate information, and (2) processing strategies which may result in inaccurate or misleading inferences.

Cognitive Limits on Information Processing. Early studies of human information processing demonstrated unambiguously that decision makers could only use limited amounts of information. Miller (1956), in a classic study, demonstrated experimentally that short term memory is limited in most people to seven "chunks" of information plus or minus two (a "chunk" being the largest single item of information recognized by the processor), and that an individual's ability to "chunk" information acts as a direct constraint on the input of data. What is important to us is that this fundamental physiological limitation appears to act as a constraint on all decision makers. Numerous studies have demonstrated that physicians, stockbrokers, meteorologists, policymakers, and a

variety of other "experts" all appear to use only a very limited number of cues in making complex decisions (e.g., Dawes & Corrigan, 1974; Slovic et al., 1977; Slovic & Lichtenstein, 1971). Other research in this area has also shown that variables such as personality, cognitive structure, and demographics are related to the ability to process information (e.g., Nystedt, 1972; Schroder, Driver, & Streufert, 1967; Taylor & Dunnette, 1974). Further evidence of limitations on the ability to process information is available in studies showing the biased nature of memory and perception. Buckhout (1974), for example, lists three common sources of unreliability in recall: (1) insignificance in the original situation, i.e., cues which were later deemed important were not closely attended to at the time, (2) the degradation of information in memory over time, and (3) pressures and distractions on the information processor which reduce the amount and accuracy of information used.

Each of these limits is potentially important for an understanding of how decision makers may use evaluation information. For instance, given that decision makers are able to use relatively little information, it becomes problematic to know which information items of the total quantity available a decision maker will focus on and use. Studies of experts are consistent in finding that experts on the same subject typically use different information in making expert judgments (e.g., Slovic et al., 1977). This suggests that decision makers, when presented with evaluation research results, may interpret and weight the information differentially. Hawkins et al. (1978), for example, in a drug evaluation study, showed that various actors were weighting information differently according to the evaluation criteria they were using. Other studies have also demonstrated variations in preferences for types of information across decision makers (e.g., Kilmann & Mitroff, 1976), as well as in how stress reduces one's ability to process information (e.g., Wright, 1974). Since, as Mintzberg (1973) has shown, managers' work is characteristically fragmented and subject to distractions and time pressures, it is likely that users of evaluation research will be unable to assimilate fully all the information contained in a report. Instead, users are likely to form overall impressions, subject to the biases mentioned previously, and to weight the results accordingly. This interpretation is consistent with studies which show that with the passage of time details are forgotten, that the reconstructed meaning is often less ambiguous than originally portrayed, and is then interpreted as offering support for a favored

position (e.g., Buckhout, 1974; Ross, 1977); that is, information contained in the original signal which is unfavorable is likely to be either forgotten or reinterpreted so as to minimize its negative consequences.

These and other studies of individual limits on cognitive information processing corroborate a reasonable but often neglected fact: memory is a selective and often fallible source of information. The perceptual process of detecting and attending to stimuli is affected by factors such as stress. With the passage of time there is also a tendency to reorganize our "memory" into coherent recollections by stripping away contradictory evidence and filling in any gaps with "constructed" facts. These limitations suggest the following propositions:

> *Proposition 7:* Only limited amounts of information from an evaluation study are likely to be used for decision-making purposes.
>
> *Proposition 8:* Given the same evaluation information, different experts will use different parts of the report in different ways; that is, judges will weight differently the same evaluation information.
>
> *Proposition 9:* Over time, decision makers will be more likely to interpret favorable information from an evaluation study as less ambiguous than originally perceived.
>
> *Proposition 10:* Over time, decision makers will be more likely to forget unfavorable information from an evaluation study or reinterpret such information as either irrelevant (e.g., does not address the "relevant" question) or favorable.

Selective Processing. Aside from these limits on decision makers' ability to process information, there also exist biases in the manner in which information is cognitively processed by individuals. Three of these biases have direct ramifications for evaluation research utilization: (1) selective perception, (2) self-serving biases, and (3) a preference for vivid, concrete information for use in decision making. The first relates to the tendency, described in the previous section, for users to "reinterpret" acquired information to fit preconceptions or to allow the user to maintain a consistent set of attitudes and beliefs about a given topic. This bias toward consistency injected by the reconstruction of facts is dramatically increased through the three selective processing mechanisms mentioned above.

Janis and Mann (1977) offer a number of excellent illustrations of instances in which decision makers either defensively avoided acquiring or processing unfavorable information or bolstered their position through the selective acquisition and interpretation of favorable data. This tendency may be seen in decision makers' willingness to overweight negative information when they desire to make a negative decision (Kanouse & Hansen, 1972). In these circumstances, decision makers wishing to reject an opposing view have been shown to use whatever negative information is available. Selection interviewers, when presented with a large number of positive cues and very few negative ones, have been shown to attend to the negative information systematically and to use it to reject applicants even though the positive information is far more potent. Miller and Rowe (1967), for example, found that when subjects were required to make assessment decisions, there was a significant tendency to be influenced by negative rather than positive adjectives used to describe a candidate. Other corroborative evidence is available from studies of personal perception among those who make investment decisions, among gamblers, and others (Kanouse & Hansen, 1972). When decision makers favor a position, the bias has been shown to operate toward the selection of favorable information as well (e.g., Morlock, 1967).

It should be noted that this bias does not necessarily suggest that decision makers truncate their search for information once having obtained data which can be used to support a desired position or oppose an undesired one. The apparent tendency is to selectively seek out information which bolsters one's position and avoid unsupportive information in either acquisition or processing, but not necessarily to avoid searching. In fact, a number of laboratory studies have demonstrated an interesting propensity among decision makers to desire more information than can be effectively used (Chervany & Dickson, 1974). The paradox is that decision makers appear to seek more information than required, even to the point of inducing overload. While the overload may actually impair performance, the additional information has been shown to increase the decision maker's confidence (Chervany & Dickson, 1974; Oskamp, 1965). The net results may be that decision makers arrive at poorer decisions but are more confident in their choices.

Thus, it may be that decision makers will selectively seek out information which supports or opposes a position, acquire as much of this information as possible, and be increasingly confident in

their decision although such decisions may be substantially biased. Meltsner (1976), in a book on policy analysts in bureaucratic settings, makes a relevant distinction between two categories of information sought by decision makers; that is, information used to *make* decisions and information used to *support* decisions. The latter category is indicative of the type sought by decision makers to justify a position. Meltsner describes at some length how it is not uncommon for decision makers to hire outside consulting groups to do evaluation studies, not to be used for decision-making purposes but solely to provide credible information which supports a decision which has already been made. This idea is directly related to the earlier discussion of information as a political resource and suggests the following proposition:

> *Proposition 11:* Results of evaluation research are more likely to be used if portions of the study can be selectively interpreted as either supportive of a desired position or unsupportive of opposing positions.

Self-Serving Biases. Aside from propensities to perceive and process information selectively, decision makers have also been shown to engage consistently in what has been labelled "self-serving" biases; that is, researchers have noted that individuals often view themselves more favorably than seems objectively warranted (e.g., Miller & Ross, 1975). Thus, for example, bettors have been shown to overpredict their gambling successes consistently (Blasovich, Ginsburg, & Howe, 1975), production managers to overpredict their performance (Kidd & Morgan, 1969), and corporate presidents to overpredict their firm's success in meeting competition (Larwood & Whitaker, 1977). Aside from this future-oriented optimism, investigators have also shown that, in retrospect, members of successful groups see themselves as more responsible for their groups than do members of groups that have failed. Schlenker and Miller (1977), for example, found that members of groups that had failed assigned less responsibility for the group's poor performance to themselves than they typically assigned to any other member of the group.

In organizations, such pervasive "self-serving" biases have important ramifications for the acquisition and interpretation of information. These biases become especially important when participants are linked to, or are responsible for, previous decisions to allocate resources. Under these conditions, decision makers may

be committed to a particular program. Staw and his colleagues (e.g., Staw, 1974; Staw & Fox, 1977) have demonstrated how commitment to a course of action may result in escalating commitments of resources to failing projects. Staw and Ross (1978), for instance, demonstrated that policymakers who had allocated resources to projects which subsequently failed for reasons that they should have foreseen, were more likely to devote even more resources to the project in succeeding time periods than were decision makers who had sponsored successful projects or whose projects had failed for exogenous reasons beyond the decision maker's control. This example is similar to foreign policy failures such as the United States' involvement in Vietnam. Staw and Ross quote George Ball (1965), who in the early years of the Vietnam War stated that, "Once we suffer large casualties . . . our involvement will be so great that we cannot—without national humiliation—stop short of achieving our complete objectives." These examples are disturbingly similar to certain educational programs which persist in spite of evaluation research which documents their failure to achieve stated objectives.

Two possible reasons for such events can be offered. First, the stated objectives of the program which are evaluated and found wanting are not representative of either the true objectives of the program (for example, when a program is established for political purposes and evaluated on educational attainments), or the evaluation is done on only a subset of the total program goals. Or, second, evaluation information attesting to the program's inadequacies is ignored. It is in this latter instance that the effects of commitment described by Staw may be linked with self-serving biases.

It may be that the ability of individuals to be over-optimistic about future events predisposes them to commit themselves to courses of action. Further, when cohesive groups are involved, there may be an even greater impetus to choose risky options. Once committed, the selective perception biases described earlier can act to provide information supportive of the original decision. Halberstam (1972) provides numerous illustrations of how Robert McNamara and others engaged in this activity during the war in Vietnam. Gouran (1976) provides similar examples showing that Nixon and his aides persistently discounted the importance of evidence during the Watergate cover-up. Selective perception biases allow the parties involved to choose information, as suggested in Proposition 11, which either supports the aims of the programs or

rebuts opposition claims. Self-serving biases may also allow for the development of a false consensus or the illusion that one's behaviors and choices are common and appropriate while opposing responses are uncommon and not widely supported. These self-serving biases also act to make the decision maker reluctant to abandon a chosen course of action. As long as information is available which can be interpreted as supportive of a given position, the bias on the part of the central actor will be to focus on this corroborative information. As Pfeffer and Salancik (1977) have shown, when ambiguity exists, particularistic criteria can be used by decision makers; that is, unless a widespread consensus exists, it is possible for opposing decision makers to argue for their positions and to selectively use information to support their claims. Given that evaluation is often prescribed precisely because the situation is ambiguous with respect to a given program, it is obviously the case that the selective processing of information will occur. Under these circumstances, individuals are likely to be involved and committed to particular points of view and self-serving biases will be operating. This impact may be heightened since the results of evaluation research may also be ambiguous.

A final effect of self-serving biases may be seen should a program be declared a failure, either for political or substantive reasons. Under these circumstances, self-serving biases may allow involved participants to cognitively reconstruct their involvement and devalue their responsibility for the failure; selective perception may act to focus on exogenous events which explain lack of success in terms of others' actions and unforeseeable events. Thus, the operation of self-serving biases suggests the following propositions relevant to an understanding of the use of evaluation information.

Proposition 12: Responsibility for a program results in increased commitment on the part of decision makers and increases the likelihood that the results of evaluation research will be used when:

a. the evaluation information can be interpreted as favorable or supportive of the program.
b. the evaluation information is ambiguous and can be argued as not reflective of the overall scope of the program.
c. no strong consensus exists as to the specific goals of the program such that some evaluation information may be interpreted as favorable to some parts of the program.

Proposition 13: Results of evaluation which are unfavorable to a program are more likely to be used when:

a. there exists a set of actors with objectives contrary to those associated with the evaluated program.
b. resources are scarce, so that the competition for resources is greater.

Abstract Versus Concrete Information. Typically evaluations are conducted to assess whether a program meets its goals, is successful compared to other projects, or in order to provide for the feedback of information to improve performance in succeeding time periods. Underlying these reasons is the idea that evaluation information will allow decision makers to derive inferences about causal relationships (for example, between educational interventions and student achievement). An important postulate of this process holds that causal explanations will be influenced by consensus information, i.e., information concerning base rates and how a given project fares compared to the base rate. Attribution theory researchers (e.g., Nisbett, Borgida, Crandall, & Reed, 1976) have drawn attention to the substantial amount of evidence which has failed to support the postulated effect of consensus information. For instance, Tversky and Kahneman (1974) have demonstrated this point by asking subjects to judge the probability that a target individual, described in a brief personality sketch, was an engineer, given:

(a) that he was drawn from a population of 70 engineers and 30 lawyers, or
(b) that he was drawn from a population of 70 lawyers and 30 engineers.

Knowledge of the population base rate for occupational categories had no effect whatever on judgments of the probability that the target individual was an engineer. Instead, subjects relied exclusively on the personality sketch and based the decision on the degree to which the description fitted the stereotypic engineer or lawyer. Numerous other examples are available which demonstrate that decision makers, and even scientists familiar with statistics, habitually ignore information about the population and draw recklessly strong inferences about the underlying population from knowledge of a very small sample. Tversky and Kahneman (1974) refer to this as the "law of small numbers."

Observers offer some reasons why even expert decision makers ignore base rate or consensus information: Kahneman and Tversky, for example, speculate that people may not know how to combine

base rate information and, therefore, ignore it. Nisbett, et al. (1976) propose that base rate information, by its very nature, is abstract and pallid, and may simply lack the force to persuade subjects to attend to it and to use it. People, it seems, are unmoved by dry, statistical data dear to the hearts of scientists and evaluators. As Bertrand Russell observed, "Popular induction depends on the emotional interest of the instances, not upon their number." Individuals respond to vivid, concrete information and ignore abstract data.

Nisbett, et al. (1976) offer several examples of this tendency. Consumers have long ignored medical advice to quit smoking and safety advice to fasten seat belts. Such appeals typically report numbers such as the probability of being a victim and are largely ignored. Yet when a highly visible and concrete incident occurs which people can focus on, results are often dramatic. For example, medical exhortation on the value of early detection of breast tumors long went virtually unheeded. Yet, the waiting lists at cancer detection clinics became months-long after television reported the mastectomies of Mrs. Rockefeller and Mrs. Ford. In an interesting experiment, Borgida and Nisbett (1977) provided prospective students with course evaluations based on ratings of students who had previously taken the courses. This information had little impact on course choices. In contrast, brief face-to-face comments about the courses had a substantial impact on course choices. Other studies have noted how vivid information which is non-diagnostic, that is unrelated to the decision to be made, may have an impact on the choice (Troutman & Shanteau, 1977).

Consider the ramifications of these biases for the use of evaluation information. The typical evaluation report is a document which relies heavily on the statistical analysis of data and variations from mean levels of performance. The essence of such a report is on base rates and variations from the mean. Complaints by consumers of such reports that the information is "not helpful," is "too dry," "relies too much on statistical analysis," or "doesn't get at the real problem" may, in fact be symptomatic of decision makers' inability to use abstract information. On the other hand, users of evaluation information often focus on a single, concrete, often dramatic, example even though the chosen example may not be representative of the larger picture. Patton (1978), in his discussion of the meanings of evaluation data, calls for evaluation reports which represent something meaningful to the identified information users, including efforts to reduce the mysticism of scientific

jargon and to increase the face validity of measures. Clearly, concrete and vivid examples which are accurate representations of the underlying results are likely to be accepted and remembered by decision makers more easily than compilations of statistics. Meltsner (1976, p. 234), reports the advice of a chief federal analyst that when writing two-page summaries of reports, it is important to "sprinkle them with juicy punch lines that will catch the readers' interest." Such pragmatic advice reflects the fact that not only are decision makers busy enough not to want to read reports, but they are also more likely to remember, and therefore more likely to use, vivid information. Another analyst reported that half his time was spent as a "rewrite man" trying to translate statistical material into a form which would be meaningful to the President and White House staff and agency heads. These observations underscore the bias people have toward concrete information and suggest the following proposition:

> *Proposition 14:* Results of evaluation research are more likely to be used when vivid, concrete illustrations of the conclusions are available.

SUMMARY AND CONCLUSIONS

This treatment of organizational decision makers' utilization or non-utilization of evaluation information began with a simplified model of decision making whose purpose was to highlight the importance and potential impacts of information in the decision process. I then argued that, in organizational settings, rather than following a decision-making process which proceeds from a problem to alternatives to a choice which optimizes, organizational decision makers have strong preferences for certain outcomes and act, in their information search and processing, in ways calculated to maximize the attainment of desired ends. This view recognizes that information is only one commodity which may help or hinder goal attainment. This view also recognizes the fact that some choices are unacceptable to certain decision makers, regardless of the net benefit to the larger collective. Knowledge does not necessarily equate with action.

Given the political process through which goals and objectives are negotiated among groups of organizational participants (e.g., Pfeffer & Salancik, 1978), universal agreement on any allocation of resources is unlikely. This lack of consensus makes the process

of organizational decision making a political one, often characterized by conflict and disagreement. In order to achieve a semblance of rationality, if only to satisfy constituencies outside the organization, I argued that procedures were established which, though they give the appearance of comprehensive rationality, may be more symbolic than real. Within the bounds of this "procedural rationality" I argued that individual decision makers were "contextually rational," that is, attempted to maximize goal attainment given a set of situational, organizational, and individual constraints. Thus, it may be that the requirement for evaluation is a manifestation of the need for procedural rationality in public management while the actual use of evaluation information is subject to the contextual rationality of relevant decision makers. That some evaluations are conducted as *pro forma* arrangements with funding agencies to insure compliance with regulations, but not intended for use in decision making, can be interpreted as an example of this procedural rationality. The fact that other evaluations are conducted to support previously made decisions may be an example of the contextual rationality of decision makers.

Since the decision-making process as illustrated in Figure 1 is an interactive one, and since the argument is that much of the information manipulation stems from the preferred set of outcomes, it is important to consider how these preferences are developed as well as how information is used. I proposed that two primary sets of constraints were relevant. First, in organizational settings, decision makers are seldom indifferent about outcomes. Rather, goals are assigned, for example, profit or cost margins, and control systems established to monitor and sanction responsible individuals. Power, or the ability to induce other groups or individuals to behave in prescribed ways, becomes an important consideration for goal attainment. The effect of these variables on decision makers is to make both salient and desirable a limited set of outcomes. These constraints, when coupled with potential loss of personal rewards such as status, promotion, social approval, and money, act to commit decision makers to certain outcomes.

I argued that decision makers, once committed, were then potentially subject to biases in both the acquisition and processing of information for use in decision making. Evidence was cited which demonstrated that decision makers were biased in their search for information, preferring accessible information which supported their preferences rather than contrary information, even if such information was of higher quality. Further, evidence was

also available which documented how commitment to certain desired outcomes was associated with the distortion of information in organizational communication. Hence, it may also be that the information available to unbiased decision makers may, if it has been transmitted through an organizational hierarchy, already contain inaccuracies or distortions.

The information processing of decision makers was also considered as a potential source of non-utilization of evaluation information. In pursuing desired outcomes, decision makers were often shown to perceive and interpret information selectively. It has also been shown that human information processors do not deal well with dry, statistical data, but prefer more vivid, concrete examples, even though such information may be inaccurate or misleading. These biases may be important since evaluation information is typically quantitative and statistical. It was argued that the combination of decision makers' selective perceptions of supportive information and general preference for vivid examples biases them away from the use of evaluation information unless such information is supportive.

The joint effects of the situational and individual constraints on information use by decision makers is outlined in Figure 2. Context variables such as incentive systems, group norms, and organizational structure may act to affect the information which is available to a decision maker. Context variables, as well as the manner in which information is processed cognitively, may also act to affect individual preferences for certain types of information. In turn, these variables may determine how and what information is used by decision makers.

When evaluation research is considered from the perspective developed in this chapter, and subject to the constraints presented in Figure 2, several observations about utilization of evaluation information are noteworthy. First, evaluation research, regardless of the rigor with which the study was conducted, is not likely to be regarded by decision makers as objective, nonpartisan information. Rather, such information will likely be viewed as useful to some interested parties, threatening to some, and irrelevant to others. The utilization of such information in decision making will probably reflect not any objective measure of quality of the research, but a number of factors independent of the evaluation study such as the degree of consensus or conflict among those involved in the decision process, the relative power of the participants, pressures on the primary decision makers, availability of other information,

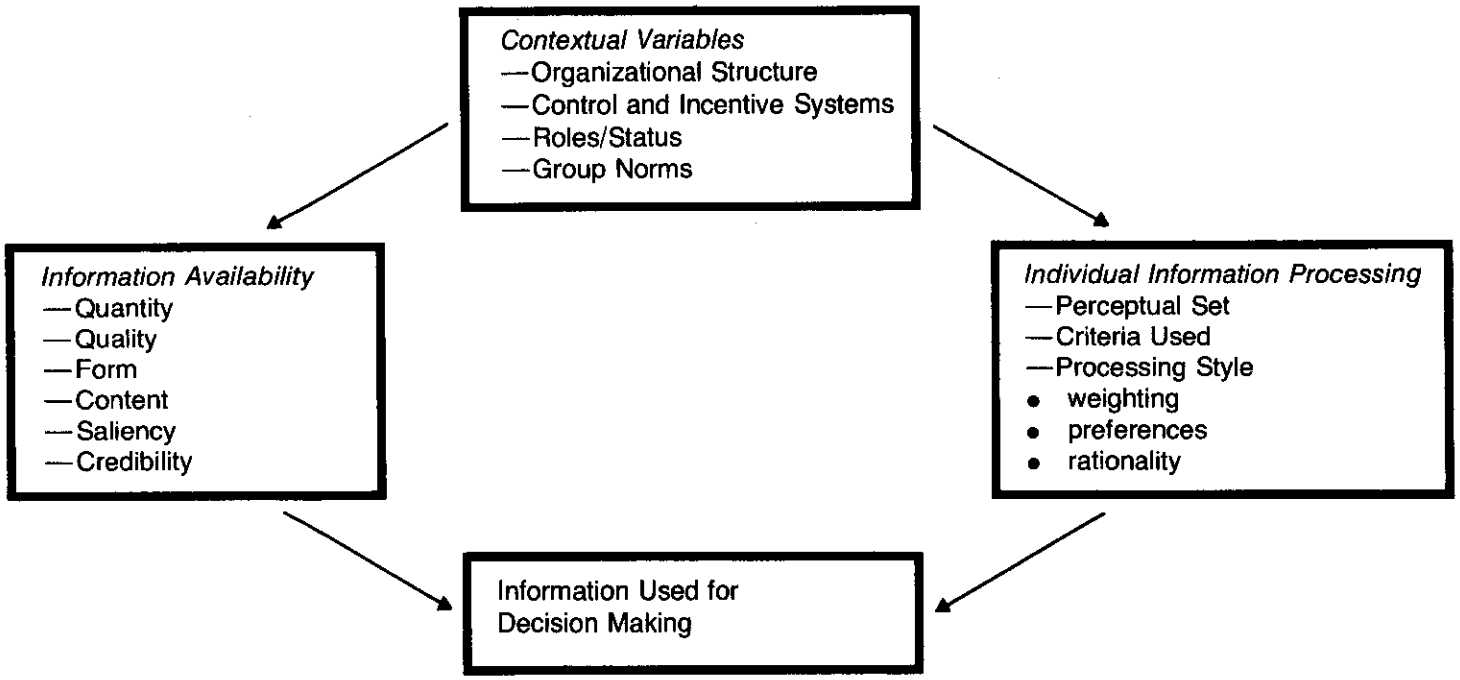

Figure 2

Contextual and Individual Variables Affecting the Use of Information
by Organizational Decision Makers

etc. In some cases, individual characteristics of the decision maker may also determine the utilization/non-utilization of evaluation results. When studies are undertaken, not to provide information for decisions, but as a means of increasing confidence in a position or for symbolic reasons, it is unlikely that any direct impact of the research will be observable. This does not imply that such information is not useful, only that its function is not directly related to decision making.

Finally, the propositions developed in this chapter are clearly tentative and somewhat simplistic and only suggest possible hypotheses which might be tested empirically. Obviously, there are a large number of influences on the use of evaluation research which have been omitted here. Therefore, what is proposed in this chapter is not a well-articulated theory of utilization of evaluation in decision making, but some tentative propositions based on previous research on organizational decision making and information use.

Future Research

Based on the empirical and theoretical evidence reviewed here, several tentative recommendations appear warranted. First, as suggested by David (1978), efforts to improve the quality of data are not likely to affect local use of evaluation. Instead, local evaluators might, as suggested in Figure 2, be attentive to understanding the goals and constraints operating on local decision makers. Rather than presuming that evaluation data are neutral and decision makers are rational, attempts might be made to present data in ways in which decision makers are likely to feel least threatened. These might include strategies to reduce commitments to competing goals, to increase trust by receivers, and to target data to specific goals.

Actions such as these may improve the utilization of evaluation information by decision makers. One may question, however, whether this effort is desirable. As David (1978) discovered, one of the primary functions of local evaluation units is to meet external reporting requirements, not to serve as a guide for program improvement. Other sources of information may be more useful in improving local programs. Perhaps local evaluation units should be serving more as a data collection and evaluation center for various interest groups as well as being involved with assessing student achievement; i.e., more responsive in terms of the provision

of information than with evaluation per se. Certainly the evidence reviewed here would suggest that there are many "good" reasons why interested decision makers might not accept information from a local evaluation agency. If these evaluation units are to have an effect on decision making and program revision, more attention needs to be paid to the interested parties' constraints and biases.

References

Alkin, M. Evaluation: Who needs it? Who cares? *Studies in Educational Evaluation*, 1975, *1*, 201–211.

Allison, G. *Essence of decision*. Boston: Little-Brown, 1971.

Ball, G. Memo to Lyndon Johnson dated July, 1965. As reported in N. Sheehan (Ed.), *Pentagon papers*. New York: Bantam, 1971.

Baum, J., & Youngblood, S. Impact of an organizational control policy on absenteeism, performance and satisfaction. *Journal of Applied Psychology*, 1975, *60*, 688–694.

Beach, L. R., Mitchell, T. R., Deaton, M. D., & Prothero, J. Information relevance, content and source credibility in the revision of opinions. *Organizational Behavior and Human Performance*, 1978, *21*, 1–16.

Bear, G., & Hodun, A. Implicational principles and the cognition of confirmatory, contradictory, incomplete and irrelevant information. *Journal of Personality and Social Psychology*, 1975, *32*, 594–604.

Blasovich, J., Ginsburg, G., & Howe, R. Blackjack and the risky shift II: Monetary stakes. *Journal of Experimental Social Psychology*, 1975, *11*, 224–232.

Blau, P. *Exchange and power in social life*. New York: Wiley, 1964.

Borgatta, E. Research problems in evaluation of health service demonstrations. *Milbank Memorial Fund Quarterly*, 1966, *44*, Part II.

Borgida, E., & Nisbett, R. E. The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology*, 1977, *7*, 258–271.

Buckhout, R. Eyewitness testimony. *Scientific American*, 1974, *231*, 23–31.

Campbell, D. T. Systematic error on the part of human links in communication systems. *Information and Control*, 1958, *1*, 334–369.

Campbell, J. P. Decision making, conflict resolution and the steady state. Unpublished manuscript. University of Minnesota, 1976.

Carter, R. Client's resistance to negative findings and the latent conservative function of evaluation studies. *American Sociologist*, 1971, *6*, 118–124.

Chervany, N., & Dickson, G. An experimental evaluation of information overload in a production environment. *Management Science*, 1974, *20*, 1335–1344.

Clausen, A. *How congressmen decide: A policy focus*. New York: St. Martin's Press, 1973.

Coleman, J. *Policy research in the social sciences*. Morristown, N.J.: General Learning Press, 1972.

Cohen, M., March J., & Olsen, J. A garbage can model of organizational choice. *Administrative Science Quarterly*, 1972, *17*, 1–25.

Connolly, T. Information processing and decision making in organizations. In B. Staw & G. Salancik (Eds.), *New directions in organizational behavior*. Chicago: St. Clair, 1977. Pp. 205–234.

Cox, G. Managerial style: Implications for the utilization of program evaluation information. *Evaluation Quarterly*, 1977, *1*, 499–507.

David, J. *Local uses of Title I evaluations*. Stanford Research Institute Report EPRC 21, July, 1978.

Davis, H., & Salasin, S. The utilization of evaluation. In E. Struening & M. Guttentag (Eds.), *Handbook of evaluation research*. (Vol. 1). Beverly Hills: Sage, 1975. Pp. 621–666.

Dawes, R., & Corrigan, B. Linear models in decision making. *Psychological Bulletin*, 1974, *81*, 95–106.

Dewhirst, H. D. Influence of perceived information-sharing norms on communication channel utilization. *Academy of Management Journal*, 1971, *14*, 305–315.

Erez, M. Feedback: A necessary condition for the goal setting-performance relationship. *Journal of Applied Psychology*, 1977, *62*, 624–627.

Etzioni, A. Mixed scanning: A 'third' approach to decision-making. *Public Administration Review*, 1967, *27*, 385–392.

Floden, R., & Weiner, S. Rationality to ritual: The multiple roles of evaluation in governmental process. *Policy Sciences*, 1978, *9*, 9–18.

Gershuny, J. Policymaking rationality: A reformulation. *Policy Sciences*, 1978, *9*, 295–316.

Gerstberger, P., & Allen, T. Criteria used by research and development engineers in the selection of an information source. *Journal of Applied Psychology*, 1968, *52*, 272–279.

Giffin, K. The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. *Psychological Bulletin*, 1967, *68*, 104–120.

Goldstein, M., Marcus, A., & Rausch, N. The nonutilization of evaluation research. *Pacific Sociological Review*, 1978, *21*, 21–44.

Gordon, G., & Morse, E. Evaluation research. *Annual Review of Sociology*, 1975, *2*, 339–361.

Gouran, D. The Watergate cover-up: Its dynamics and its implications. *Communications Monographs*, 1976, *43*, 176–186.

Halberstam, D. *The best and the brightest*. New York: Random House, 1972.

Harrell, A. The decision-making behavior of air force officers and the management control process. *Accounting Review*, 1977, *52*, 833–841.

Hawkins, J., Roffman, R., & Osborne, P. Decision makers' judgements: The influence of role, evaluative criteria and information access. *Evaluation Quarterly*, 1978, *2*, 435–454.

Hickson, D., Hinings, C., Lee, C., Schneck, R., & Pennings, J. A strategic contingencies' theory of intraorganizational power. *Administrative Science Quarterly*, 1971, *16*, 216–229.

Janis, I. *Victims of groupthink*. Boston: Houghton-Mifflin, 1972.

Janis, I., & Mann, I. *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: Free Press, 1977.

Janowitz, M., & Delany, W. The bureaucrat and the public: A study of information perspectives. *Administrative Science Quarterly*, 1957, *2*, 141–142.

Johnson, R. J. Conflict avoidance through acceptable decisions. *Human Relations*, 1974, *27*, 71–82.

Kanouse, D. W., & Hanson, L. R. Negativity in evaluations. In E. E. Jones, et al. (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1972. Pp. 47–62.

Kaufman, H. *Administrative feedback.* Washington, D.C.: Brookings, 1973.
Kerr, S. On the folly of rewarding A, while hoping for B. *Academy of Management Journal,* 1975, *18,* 769–783.
Kidd, J., & Morgan, J. A predictive information system for management. *Operational Research Quarterly,* 1969, *20,* 149–170.
Kilmann, R., & Mitroff, I. Qualitative versus quantitative analysis for management science: Different forms for different psychological types. *Interfaces,* 1976, *6,* 17–27.
Larwood, L., & Whittaker, W. Managerial myopia: Self-serving biases in organizational planning. *Journal of Applied Psychology,* 1977, *62,* 194–198.
Latham, G., & Yukl, G. A review of research on the application of goal setting in organizations. *Academy of Management Journal,* 1975, *18,* 824–845.
Levine, A., & Levine, M. The social context of evaluation research. *Evaluation Quarterly,* 1977, *1,* 515–542.
Lindblom, C. E. The science of 'muddling through.' *Public Administration Review,* 1959, *19,* 79–88.
Lowe, E., & Shaw, R. An analysis of managerial biasing: Evidence from a company's budgeting process. *Journal of Management Studies,* 1968, *5,* 304–315.
Lyon, C. D., Doscher, L., McGranahan, P., & Williams, R. *Evaluation and school districts.* Center for the Study of Evaluation, University of California, Los Angeles, 1978.
MacCrimmon, K. R. Managerial decision-making. In J. W. McGuire (Ed.), *Contemporary management: Issues and viewpoints.* Englewood Cliffs, N.J.: Prentice-Hall, 1974. Pp. 445–495.
March, J. G. Bounded rationality, ambiguity, and the engineering of choice. *Bell Journal of Economics,* 1978, *9,* 587–608.
March, J. G., & Simon, H. *Organizations.* New York: Wiley, 1958.
McCleary, R. How parole officers use records. *Social Problems,* 1977, *24,* 576–589.
Meltsner, A. J. *Policy analysts in the bureaucracy.* Berkeley: University of California Press, 1976.
Menzel, H., & Katz, E. Social relations and innovation in the medical profession. *Public Opinion Quarterly,* 1955, *19,* 337–352.
Meyer, J., & Rowan, B. Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology,* 1977, *83,* 304–363.
Miller, G. A. The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 1956, *64,* 81–97.
Miller, D. T., & Ross, M. Self-serving biases in the attribution of causality: Fact or fiction. *Psychological Bulletin,* 1975, *82,* 213–225.
Miller, J. W., & Rowe, P. M. Influence of favorable and unfavorable information upon assessment decisions. *Journal of Applied Psychology,* 1967, *51,* 432–435.
Mintzberg, H. *The nature of managerial work.* New York: Harper and Row, 1973.
Mintzberg, H., Raisinghami, D., & Theoret, A. The structure of unstructured decision processes. *Administrative Science Quarterly,* 1976, *21,* 246–275.
Mohr, L. On organizational goals. *American Political Science Review,* 1973, *66,* 470–481.
Morlock, H. The effect of outcome desirability on information required for decisions. *Behavioral Science,* 1967, *12,* 296–300.

Nisbitt, R. E., Borgida, E., Crandall, R., & Reed, H. Popular induction: Information is not necessarily informative. In J. S. Carroll & J. W. Payne (Eds.), *Cognitions and social behavior*. Hillsdale, N.J.: Erlbaum, 1976. Pp. 113–133.

Nystedt, L. Judgements in quasi-clinical situations: Accuracy and process. Reports from the Psychological Laboratories #12, University of Stockholm, April, 1972.

O'Reilly, C., & Roberts, K. Relationships among components of credibility and communication behaviors in work units. *Journal of Applied Psychology*, 1976, *61*, 99–102.

O'Reilly, C. A. Superiors and peers as information sources, work group supportiveness, and individual decision making performance. *Journal of Applied Psychology*, 1977, *62*, 632–635.

O'Reilly, C. A. Variations in decision maker use of information sources: The impact of quality and accessibility of information. Unpublished manuscript, University of California, Berkeley, February 1979.

O'Reilly, C., & Pondy, L. Organizational communication. In S. Kerr (Ed.), *Organizational behavior*. Columbus, Ohio: Grid, 1979.

Orlans, H. *Contracting for knowledge*. San Francisco: Jossey-Bass, 1973.

Oskamp, S. Overconfidence in case study judgements. *Journal of Consulting Psychology*, 1965, *29*, 261–265.

Patton, M. W. *Utilization-focused evaluation*. Beverly Hills: Sage, 1978.

Perrow, C. *Organizational analysis: A sociological view*. Belmont, Calif.: Wadsworth, 1970.

Pettigrew, A. Information control as a power resource. *Sociology*, 1972, *6*, 187–204.

Pfeffer, J., & Salancik, G. Organizational decision making as a political process: The case of a university budget. *Administrative Science Quarterly*, 1974, *19*, 135–151.

Pfeffer, J., & Salancik, G. Administrator effectiveness: The effects of advocacy and information on achieving outcomes in an organizational context. *Human Relations*, 1977, *30*, 641–656.

Pfeffer, J., & Salancik, G. *The external control of organizations: A resource dependence perspective*. New York: Harper and Row, 1978.

Plott, C., & Levine, M. A model of agency influence on committee decisions. *American Economic Review*, 1978, *68*, 146–160.

Porat, A., & Haas, J. Information effects on decision making. *Behavioral Science*, 1969, *14*, 98–104.

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York: Academic Press, 1977. Pp. 173–220.

Salancik, G. R. Commitment and the control of organizational behavior and belief. In B. M. Staw & G. R. Salancik (Eds.), *New directions in organizational behavior*. Chicago: St. Clair, 1977. Pp. 1–54.

Schlenker, B. R., & Miller, R. S. Egocentrism in groups: Self-serving biases or logical information processing? *Journal of Personality and Social Psychology*, 1977, *35*, 755–764.

Schroder, H., Driver, M., & Streufert, S. *Human information processing*. New York: Holt, Rinehart and Winston, 1967.

Simon, H. On the concept of organizational goals. *Administrative Science Quarterly*, 1964, *9*, 1–22.

Simon, H. Rationality as a process and as product of thought. *American Economic Review*, 1978, *68*, 1–16.

Slovic, P., Fischhoff, & Lichtenstein, S. Behavioral decision theory. In M. Rozenzweig & L. Porter (Eds.), *Annual Review of Psychology*. Palo Alto: Annual Review, Inc., Vol. #28. 1977. Pp. 1–39.

Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgement. *Organizational Behavior and Human Performance*, 1971, *6*, 649–744.

Staw, B. M. Attitudinal and behavioral consequences of changing a major organizational reward: A natural field experiment. *Journal of Personality and Social Psychology*, 1974, *30*, 724–751.

Staw, B. M., & Fox, F. V. Escalation: Some determinants of commitment to a previously chosen course of action. *Human Relations*, 1977, *30*, 431–450.

Staw, B. M., & Ross, J. Commitment to a policy decision: A multi-theoretical perspective. *Administrative Science Quarterly*, 1978, *23*, 40–64.

Streufert, S. C. Effects of information relevance on decision making in complex environments. *Memory and Cognition*, 1973, *1*, 224–228.

Taylor, R., & Dunnette, M. Relative contribution of decision-maker attributes to decision processes. *Organizational Behavior and Human Performance*, 1974, *12*, 286–298.

Thompson, J. D. *Organizations in action*. New York: McGraw-Hill, 1967.

Troutman, C. M., & Shanteau, J. Inferences based on nondiagnostic information. *Organizational Behavior and Human Performance*, 1977, *19*, 43–55.

Tversky, A., & Kahneman, D. Judgement under uncertainty: Heuristics and biases. *Science*, 1974, *185*, 1124–1131.

Ward, D., & Kassebaum, G. On biting the hand that feeds. Some implications of sociological evaluations of correctional effectiveness. In C. Weiss (Ed.), *Evaluating action programs: Readings in social action research*. Boston: Allyn & Bacon, 1972. Pp. 300–310.

Weick, K. Enactment processes in organizations. In B. Staw & G. Salancik (Eds.), *New directions in organizational behavior*. Chicago: St. Clair, 1977. Pp. 267–300.

Wildavsky, A. *The politics of the budgetary process*. Boston: Little-Brown, 1974.

Wilensky, H. *Organizational intelligence*. New York: Free Press, 1967.

Winkler, R., & Murphy, A. Experiments in the laboratory and the real world. *Organizational Behavior and Human Performance*, 1973, *10*, 252–270.

Witte, E. Field research on complex decision-making processes—the phase theorem. *International Studies of Management and Organization*, 1972, 156–182.

Wright, P. The harassed decision maker: Time pressures, distractions, and the use of evidence. *Journal of Applied Psychology*, 1974, *59*, 555–561.

Notes on Zucker's "Role of Evaluation Units in Schools"

Introduction

Lynne Zucker is a sociologist at UCLA. In her chapter, she uses data from the CSE survey of school district evaluation units to examine the role that these units play in signaling to external funding sources that their parent district organization is complying with regulations and is therefore entitled to continued support. Before re-analyzing the CSE survey data, Zucker introduces several propositions about how formal structures develop in organizations such as school districts. She contrasts the emergence of these structures in school districts, which must be responsible to societal demands, with the emergence of formal structures in other institutions which are more technically or task-oriented. Her chapter builds on the conceptualizations of Weick (1976) and Meyer and Rowan (1977) who have described how the technical core activities of educational organizations are disconnected from or only loosely coupled with formal administrative control structures.

Argument

Zucker posits, in agreement with the authors cited above, that schools and school districts belong to a class of organizations regarded as essential in achieving societal goals. Because of this, many of their internal operating units are initiated and maintained primarily to respond to external expectations or requirements. While evaluation units may sometimes serve the internal operating needs of the organization, their work is frequently irrelevant or tangential to the organization's technical tasks.

Zucker notes that societal demands on school districts, in the form of federal and state laws and regulations, have increased in recent years partly because of the growth of federal and state financial support to local educational agencies. Such financial sup-

port has been distributed in accordance with categorical program guidelines and continued in light of evidence about the program's effectiveness.

To produce that evidence, many school districts have organized separate units to do evaluation and testing. Zucker contends that these evaluation units perform the signaling function alluded to earlier: that is, they signal to groups or agencies in the environment that the district is in fact complying with legislative or administrative requirements.

Zucker defends both the legitimacy of the demands made by federal and state agencies on school districts and the rationality of school districts in de-coupling the evaluation units, which respond to those requirements, from other sub-units within the district. She points out that evaluation units often collect data to tell the outside world that the district is performing adequately. Such adequate performance is the justification for additional resources to the district.

This thesis has a number of implications for how personnel in evaluation units may regard instruction and for how teachers and administrators responsible for instruction may regard evaluation units. Since evaluation units are structures which fulfill the district's need to be externally accountable their staff tend to have minimal influence over internal organizational policies. They tend to emphasize to themselves and to others the ambiguity and uncertainty inherent in the technical—that is, instructional—activities of the district. As they self-protectively, and perhaps accurately, assert their own lack of control and their own uncertainty about causality in the instructional process, they tend to produce information of little internal use. Therefore, the status and power of the unit within the central office tends to be minimal.

In re-analyzing the CSE survey data in terms of this framework, Zucker defined a number of variables related to the characteristics of the district, the characteristics of the evaluation unit, the external role of the unit, and the internal role of the unit. In support of her propositions, she found, for example, that fewer than half of the units had internal roles in the selection of curricula or program materials, allocation of funds, facilities planning, collective bargaining, or teacher performance review.

Contribution and Importance

There are many in the evaluation community who may be reluctant to accept the role of evaluation units as mere conveyors of test data or evaluation data to outside agencies. However, studies such as those of David (1978), Alkin, Daillak, and White (1979), and Daillak (1981), corroborate Zucker's view that much evaluation activity is regarded as "paperwork" by people within the system. Zucker's analysis of the CSE survey data confirmed field work suggestive of this interpretation.

The implications of these findings, however, are not totally negative. They make clear the structural dilemma in which evaluation units are placed. It may be possible for units in some districts to differentiate their external and internal functions and intentionally separate these functions from one another. It may also be that as regulations and demands for compliance diminish in the coming years because of reconceptualization of federal and state involvement in local educational agencies, the evaluation and testing capacities developed within these units can be turned to productive internal use. Or it may be that districts, once they have satisfied the signaling function, can turn their attention to serving internal organizational needs.

References

Alkin, M. C., Daillak, R., & White, P. *Using evaluations: Does evaluation make a difference?* Beverly Hills, California: Sage, 1979.

Daillak, R., Evaluators at work: A field study. CSE Report No. 154, Center for the Study of Evaluation, University of California, Los Angeles, 1981.

David, J. L. *Local uses of Title I evaluations.* Research Report EPRC 21. Educational Policy Research Center. Stanford Research Institute, 1978.

Meyer, J. W., & Rowan, B. Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology,* 1977, *83,* 340–363.

Weick, K. E. Educational organizations as loosely coupled systems. *Administrative Science Quarterly,* 1976, *21,* 1–19.

Institutional Structure and Organizational Processes: The Role of Evaluation Units in Schools[1]

Lynne G. Zucker
University of California, Los Angeles

Historically, there has been a gradual but distinct trend toward the vesting of more control in the federal and state government. Localities have not surrendered their autonomy, but have become increasingly embedded in a multi-tier system, where state and federal control and funding of local organizations has risen substantially in the past twenty years (see Zucker, 1980). While this trend is apparent in education, local school districts have been more successful than most public organizations in retaining their autonomy (Kirst, 1970; Meyer, 1979).

Schools are crucial to the pursuit of central societal goals, and therefore have been subjected to institutional definitions of what is proper educational procedure and practice. At the federal, state, and local level, school performance is monitored, evaluated, and subjected to overall assessments. With the implementation of Title I of the Elementary and Secondary Education Act of 1965, local school districts were required to evaluate their performance. Further pressure for evaluation is rooted in the increasingly heavy dependence, especially in some states, on other types of extra-local funding for school districts. These sources of funding increase the legitimacy of external demands for assessment; local district needs for evaluation to improve program performance have largely been

[1]I wish to thank Pamela S. Tolbert for her help on the data analysis. The following all provided valuable comments on an earlier draft: John W. Meyer, Maureen McConaghy, Oscar Grusky, David O'Shea, Charles O'Reilly III, Mary Ann Millsap, and Adrianne Bank. Eva Baker and Adrianne Bank, Center for the Study of Evaluation, Graduate School of Education, UCLA, provided access to data on school evaluation units and administered the contract from the National Institute of Education which supported data collection and analysis. A grant from the Committee on Research, Academic Senate, UCLA, supported collection of additional data and necessary research assistance.

eclipsed by the need for accountability to funding sources. The current "rational view" of schools held by these outside sources poses a problem for schools because of the legitimate authority the federal and state evaluation requirements have over internal school practices. "Hyper-rationalization" (Wise, 1977a and b) could be ignored were it not for the concomitant attempt by each separate funding body to enforce this view of educational process.

Conformity to institutional rules can be seen as the schools' *raison d'être.* Many of the processes in schools involve compliance to external definitions of what schools should do, and responding to external mandates for evidence of such compliance. Especially as state and federal funds have increasingly been allocated to local public schools, these external demands for evidence have become more well defined and complex. Largely in response to these demands, local public school districts have established evaluation units. Such units are not required and are generally not funded directly by federal or state sources but the requirements for evaluation data (coupled with grant and special program evaluation requirements and heavy reporting requirements in some states) make it convenient for school districts to establish an evaluation unit. Evaluation units generally serve to centralize and simplify data collection and dissemination. It is not surprising that the number of evaluation units in school districts has literally mushroomed since 1965. About a third of all school districts with over 5,000 students have evaluation units. Fully 85 percent of the evaluation units have been established since 1965 and over half since 1970 (Lyon, 1978; Lyon, Doscher, McGranahan, & Williams, 1978).

This chapter develops a theory which emphasizes the signal nature of evaluation units. Depending on the degree to which a school district exists in an institutionalized environment, the environment serves as a source of support (including funds) for it. As a result, the environment (state and federal government) exercises legitimate control and surveillance over the district's activities and performance (for a general discussion, see Pfeffer & Salancik, 1978). From this perspective, the primary function of an evaluation unit is to produce signals from the school district to this external environment. Hence, most of the information produced by the evaluation unit is in response to accountability requirements generated at the federal and state levels. Little of the information is used within the organization (school district), though it may be disseminated within it (see David, 1978, for further sup-

port of this argument). Further, the evaluation unit generally has a limited internal function (e.g., does not affect district policies or other units), with little connection between evaluation unit staff and instructional staff. In other words, the evaluation unit serves to legitimate what the school district is doing by reporting it (generally in favorable terms) to external sources of funding and control to which the district is legitimately accountable.

A general theoretical explication of the sources and consequences of institutional organizational environments is developed in the next section. Then the concepts are applied more directly to evaluation units in school districts. Some central hypotheses are tested using data drawn from CSE's national survey of evaluation units (Lyon, Doscher, McGranahan, & Williams, 1978), combined with measures of federal and state funding and of control in school district organizations.

SOURCES AND CONSEQUENCES OF FORMAL STRUCTURE

In schools, as in other large-scale organizations, formal organizational structures arise, often not clearly or closely related to the actual activities of the organization (cf. Gouldner, 1954). This formal structure is fundamentally a blueprint for organizational activity. The organization chart, with its listing of major officers, departments, and programs, details the organization's formally defined structure. These elements are linked by rationally defined connections between activities, impersonally embodied in the organization's explicit goals.

Much of modern organization theory has been concerned with exploring the sources of formal structure in organizations (Scott, 1975). Research thus far has investigated formal structure which arises primarily from problems of coordination inherent in the core technology (e.g., Woodward, 1965; Mohr, 1971; Comstock & Scott, 1977; Billings, Klimoski, & Breaugh, 1977). In contrast, the emerging theory of institutionalization (Zucker, 1977; Meyer & Rowan, 1977) has stressed the role of social definition, of "myth," in determining organizational structure. Independent of the core technology, or "loosely coupled" to it (see Weick, 1976; Meyer & Rowan, 1977), formal structure created by social definition serves to legitimate the organization. The organization incorporates elements of structure, such as evaluation units or affirmative action officers, which reaffirm organizational conformity to the externally imposed definition of what is legitimate.

This body of work on the sources of structure, whether task oriented or institutional, has treated the origin of formal structure as largely outside of the organization's direct control: the structure is determined either by the technology or by the institutionalized environment. It is important to note that wider societal definitions are involved in either case. The application of appropriate technology is defined in terms of "state-of-the-art," and organizations are expected to adopt innovations defined as advances in technological development. Institutionalized procedures, certified professionals, and programs of action involve actual organizational activity, but at the same time incorporate societal definitions of proper practices even when conformity conflicts with efficiency criteria (Meyer & Rowan, 1977). Regardless, then, of the particular advantage or disadvantage for the task performance itself, widespread definition of a technology, procedure, or division of departments as rational and legitimate leads to organizational adoption. For example, early adopters of civil service procedures exhibit characteristics which indicate their need for more formal personnel procedures, while cities adopting these procedures later in the process do not, but rather adopt them simply on the basis of their widespread legitimacy (Tolbert & Zucker, 1980).

But the organization itself should be recognized as an important determinant of its own location in the wider environment. Organizations, whether technological or institutional, are not simply passive captives of their environments. Their role in regulating environmental effects by developing boundary maintenance functions, domain definitions, and other mechanisms of control has long been recognized in organization theory (see Thompson, 1967). Most fundamentally, however, organizations seek to construct their own environments, not simply manage preexisting environmental constraints (e.g., Pfeffer's study of organizational merger, 1972). For example, organizations may define their appropriate institutional location (Dowling & Pfeffer, 1975): Training institutes wish to define themselves as educational institutions, thereby gaining access to societal resources such as tax write-offs and G. I. Bill funds, although their functions, in fact, parallel personnel agencies more closely.

The foregoing argument can be summarized as follows:

> *Proposition 1:* Organizations seek to define their own location in the wider institutional/technological environment.

Organizations define their location along a continuum, ranging from (1) narrowly construing their activities as technical and refusing societal resources (e.g., training stipends for workers) which would broaden the societal relevance of their goals, or (2) broadly construing their activities as fundamentally societal and accepting societal resources which reinforce that view.

This statement differs from earlier work on organizations in a number of respects. Probably the most important, from an operational point of view, is that the degree of dependence on societal resources is a crucial step in defining the organization as one serving societal interests, and thus existing in an institutionalized environment. Hence, public organizations are, *a priori*, operating in institutionalized environments, though the extent to which obligations of the organization are normatively defined may vary. Organizations which are not public, but which deliver services seen as related to the public good, may also operate in institutionalized environments. Increasingly, all organizations, including profit-making firms (e.g., Lockheed and Chrysler), are being identified as central to the common good, and hence as deserving of societal support.[2]

A corollary of central importance, since organizations are bound by history and convention, is:

> *Corollary:* Once an organizational type is defined as societal or technical, other new organizations will have to demonstrate that they are not of that type before redefinition can succeed.

Social definitions of appropriate procedures and practices have, over time, the force of facts. For example, educational organizations are firmly embedded in the institutional environment; it is unlikely that an educational organization could successfully redefine its environment as technological, and so escape accreditation, certification, and other institutional requirements (Meyer & Rowan, 1978).

Consequences of Institutionally Derived Structure

Little work has focused on the consequences of formal organizational structure, whether derived from technological or institutional environments. Generally, structure derived from needs of

[2] I am grateful to Marshall Meyer for suggesting this interpretation.

core technology is presumed to have consequences largely internal to the organization: altered task organization, altered efficiency (Leifer & Huber, 1977). But structure derived from institutional requirements is thought to be de-coupled from technical activities and to affect, primarily, survival rates of organizations (Meyer & Rowan, 1977).

Generally, while organizations in the more "rational" approach to structure are seen as interpenetrated by the environment, most of these inroads are either controlled or buffered to prevent them from having any direct influence on internal organizational functioning or structure (Thompson, 1967). Typically, organizations are seen as autonomous, in most senses not directly accountable to interests located outside the organization. While private and public organizations are continuously interpenetrated by customers (e.g., purchasers of cars) or clients (e.g., students of schools), it has been noted generally that these groups do not exert much control over the respective organizations (see Hasenfeld, 1972, on people-processing organizations in general). While in principle such groups exert control because they must select the product or service, they frequently operate within limited choice ranges (one must have a car and, according to the amount which can be spent, may have a choice between two or three major competitors) or are a "captive audience" (requirements for school attendance coupled with local school concepts ensure, at best, a restricted range of choice).

External Control

The key to understanding the effects of the environment or internal organizational structure appears to be the locus, extent, and legitimacy of external control over internal organizational process. In normal market environments, organizations are viewed as legitimately autonomous in pursuit of their own goals as defined by them. In striking contrast to this, organizations in institutional environments are seen as properly accountable to societal interests, and therefore as legitimately controlled by societal agents. To put it more directly, external administrative and legislative control is thought to be desirable in order to ensure that such organizations are, indeed, serving the societal interests they were created to serve or perpetuate.

Such external control necessitates the development of rules and procedures, and also rather formal rules for evaluation. As Kaufman (1960) has pointed out, control without direct supervision (which

has been termed "long range control") necessitates development of elaborate monitoring and evaluation systems and other control techniques, such as professionalization, to maximize "self-regulation." It should be noted that all public organizations can legitimately be held accountable, since they are ostensibly organized for the public good. It is not the case, however, that all are actually so controlled. For example, when secrecy is essential (CIA, FBI) such control is relinquished. Further, some control is thought to be legitimate in private industry to the extent that the common good is potentially involved (e.g., railroads, air transportation, "vital resource" industries such as coal and steel).

A second proposition can now be stated:

> *Proposition 2:* The greater the extent to which an organization exists in an institutional environment, the greater the degree of legitimate control and demands for accountability coming from the relevant environment.

The environmental control and demands for accountability are problematic to the extent that direct surveillance of organizational performance cannot be performed. As long as control is local, direct evidence can be obtained. But when funding and control come from extra-local sources, as in the case of school districts, accountability becomes more problematic. Effectively, the organization is required to become "self-evaluating" (Wildavsky, 1972), and to transmit the information obtained to the external sources of funding and control. A third proposition, then, is:

> *Proposition 3:* If legitimate control and demands for accountability are extra-local, then organizations will be required to perform evaluation (or, minimally, monitoring) of their own activities.[3]

At the same time, however, the support and funding the organization requires may be contingent on the results of the evaluation. One major consequence of this is the creation of internal organizational uncertainties and inconsistencies. Loose coupling of administrative structure from the rest of the organization (and activities from each other) may, instead of being dysfunctional, be the most rational strategy for maximizing production of positive

[3] O'Shea discusses differences between evaluation and monitoring elsewhere in this volume.

evaluation results. Similarly, subunits responsible for responding to external demands for accountability will be loosely coupled to subunits performing other activities. This point will be developed more fully below.

Formal Structure and Signaling

Much as individuals use signals in negotiating the labor market (see Spence, 1974), organizations use signals to provide information to their environments. Organizations develop tasks, programs and goals, and even subunits to signal their intentions to the environment (on the latter point, see M. Meyer, 1980). Depending on the control which the environment has over the organization, the organization will proliferate formal structure which produces signals demonstrating accountability to the relevant environment.

In brief, thus:

> *Proposition 4:* To the extent that the environment has legitimate control over an organization, the organization will produce structure which serves a signaling function.

Even organizations which are largely autonomous from the institutional environment may, under some conditions, be sufficiently constrained by it to produce structure which serves a signaling function. A case in point is the widespread existence of antitrust departments in corporations. However, organizations more deeply embedded in the institutional environment (Zucker, 1977) have elaborated many aspects of formal structure which produce signals demonstrating accountability—evaluation units in schools, designed to transmit information concerning the quality and degree of task focus, and "management information systems" (MIS units) in local employment and training programs, designed to transmit evidence of internal competence in processing clients.

Formal structure which serves signaling functions is, in many respects, equivalent to boundary personnel: It represents the organization to the wider environment, signals its adequate performance as a means of generating additional resources, but has little internal responsibility (see Zucker, 1979).[4] Formal structure

[4] Its responsibility can be defined as "environmental management," with few internal responsibilities.

which serves largely boundary functions appears loosely coupled to the rest of the organizational structure precisely because it is not directly responsible for, or not focused on, internal task performance. Hence:

> *Proposition 5:* To the extent that components of formal structure are designed to send signals to the wider institutional environment, they are loosely coupled to internal organizational performance.

Institutional Structure and Evaluation Criteria:

What, then, of the content of the formal structure and the activity it generates? First, the content of formal structure created as a response to non-institutional, technological environments is primarily internal and task-focused, while that created as a response to institutional environments is primarily externally focused. Hence,

> *Proposition 6:* Formal structure generated in response to technological contingencies is concerned with internal organizational functioning, while formal structure generated in response to institutional contingencies is focused on external representation of organizational functioning.

Second, the actual activities generated by the formal structure will be different: Organizations operating in institutional environments, as schools do, define their scope of activity to require being less and less certain about more and more (emphasizing ambiguity and uncertainty), while organizations not so externally accountable define their scope of activity to require being more and more certain about less and less (emphasizing control and certainty). Schools (and similar organizations) need to emphasize the uncertainty/ambiguity because they are held externally accountable to different (and changing) standards of what is institutionally proper. Business organizations are not held accountable externally to the same degree and the external standards which do apply are more uniform (e.g., anti-trust regulations).

Further, the greater the interpenetration of the local organizations by institutional demands and resources, the more ambiguous and uncertain activities are presented:

> *Proposition 7:* The greater the extent to which an organization exists in an institutional environment, and the external evaluation therefore becomes more critical, the more the organization stresses the

ambiguity of evaluation criteria in order to make external evaluation more problematic.

An important consequence of ambiguous criteria is that the organization becomes less able to evaluate its own performance. For example, before schools were heavily dependent on external (extra-local) support, validation, and resources, they presented themselves as producing rather specific training, evaluated on grounds of efficiency; as schools moved into a more institutionalized environment, they defined their tasks as increasingly ambiguous and uncertain, best evaluated on grounds of certification (see Meyer & Rowan, 1975; compare Callahan, 1962 to Tyack, 1974).

Organizations in institutional environments use the evaluation of task performance as a signal, alterable and ambiguous, to indicate compliance to institutional rules, not to guide adjustments in task organization. Organizations in technological environments use evaluation of task performance as an index of how well they are performing, as clear, unalterable evidence which can be used to guide adjustments in task organization.[5] Characteristics of the task itself determine the extent to which it can be redefined as an index or a signal. Teaching, for example, is more ambiguous than automobile assembly (for a general discussion of active and inert tasks, see Dornbusch & Scott, 1975). Basically, it is asserted that organizations in institutional environments will attempt to define their task performance in a different way than organizations not held accountable to societal interests. Fundamentally:

(1) Organizations will act to produce and/or apply largely ambiguous criteria for task performance when societal support and funding depend on evaluation:
   (a) They will define the environment as more variable, increasing the scope of activities and making more aspects of the environment appear relevant to their tasks (e.g., range of pupil characteristics);

[5] Spence (1974) assumes that education is a signal, that is, an alterable observable characteristic, while race is an index, that is, an unalterable observable characteristic. However, sociologists would argue that education will in most instances serve as an index (though affected by individual choice, the employer in fact assumes it is fixed at the current level), while race is not uniformly an index, as the phenomenon of "passing" clearly illustrates. However, the willingness to identify and rely on signals or indices is of interest here.

(b) They will define tasks as more ambiguous and difficult (e.g., how can "good" teaching be defined, let alone regulated?); and
(c) They will use evaluation information primarily as signals to the external environment.

(2) Organizations act to reduce, and/or render inapplicable, institutional rules when evaluation control along with resources are centered in the local organizational units (or when resources are controlled by a large and diffuse "public"):
(a) They will define the relevant environment to make it seem more constant by reducing the scope of activities and range of relevance (e.g., only serve one small segment of the market);
(b) They will define tasks as clear, limiting the scope of the task (e.g., not to produce a "good" product, but one which meets certain pre-set specifications); and
(c) They will use evaluation information primarily internally (it may, in fact, be secret or privileged), as indices of task performance. Finally, the status of the organizational members of units responsible for evaluation will vary, depending on the degree to which evaluation serves a signaling function. Thus:

Proposition 8: The greater the signaling function of evaluation, and hence, the more it is viewed as ambiguous, the lower the internal status of the members or units responsible for evaluation.

When evaluation simply serves to legitimate an organization, and is defined as largely ambiguous and of little value to the organizational task performance, then the members or units which produce it will be given minimal resources and support.

In the next section, the relevance of this general model for evaluation units in school districts is established.

THE ROLE OF EVALUATION UNITS IN SCHOOL DISTRICTS

It is clear that educational organizations do not operate in the rational technology-oriented fashion that most organizational theorists have assumed (see especially Meyer, 1977; Meyer & Rowan, 1977, 1978). Separate organizational subcomponents and levels of authority do not articulate well with one another but,

instead, operate relatively autonomously with little control exercised (Weick, 1976 and Cohen, March, & Olsen, 1972, provide good discussions of these issues). The lack of control over central task processes has been labeled loose coupling. The primary reason for the presence of loose coupling in educational institutions appears to be that they function in environments which demand conformity to institutional rules rather than to technological perfection or innovation, with some subunits largely engaged in producing signals which indicate to the wider environment that the school is accountable to societal interests. As Meyer (1977, pp. 4–5) puts it,

> . . . the technical organization faces in toward its technical core and turns its back on the environment; the institutional organization turns its back on its technical core in order to concentrate on its institutional environment . . . [the] crucial thing a school needs to do to survive is to conform to institutional rules—including community understandings—defining teacher categories and credentials, pupil selection and definition, proper topics of instruction, and appropriate facilities.

District-wide evaluation units have appeared, largely as a response to increased funding and control at the state and federal level, to handle reporting, monitoring, and evaluation requirements. On a nationwide basis, 85 percent of all evaluation units appeared following the implementation of federal evaluation requirements accompanying Title I in 1965 (David, 1978). In the CSE survey, only 3 out of 259 districts (1.2%) reported that they have no state or federally funded programs. State or federally funded programs are evaluated by the evaluation unit in over 80 percent of the districts, while in less than 14 percent is this task delegated to other district units or personnel (and in less than 5% delegated to outside evaluators solely).

The direct relationship between the resources coming into the district and control is indicated in the following quote from a director of a Title I program (David, 1978, p. 13): "This district will accept all the strings that go with the Federal money." There is little doubt that the primary function the evaluations serve is to meet state and federal reporting requirements. As one evaluator explicitly states, evaluation is generally associated with accountability, not with information useful in assessing strengths and weaknesses of programs (David, 1978, p. 39): "I don't know whether

the test scores are useful as a basis for making changes in the program because I don't deal with the content of the program."

Before beginning a more extensive analysis of the data on evaluation units collected in the CSE survey, the central hypotheses, and their relationship to the propositions above, need to be stated explicitly. First, as the percent of societal resources increases, the amount of regulation and control increases (Proposition 2). Second, as these resources are increasingly extra-local, the organization will proliferate formal structure (units, positions) responsible for self-evaluation (Proposition 3). Third, as the degree of institutionalization and extra-local control increases, the percent of self-evaluation information used externally also increases (the signal nature of evaluation, from Propositions 4 and 6). Fourth, as the information is increasingly used externally, the unit/personnel responsible for evaluation increasingly have minimal influence over internal organizational policies, and have little connection to actual organizational tasks (loose coupling of unit/personnel from Proposition 5). Fifth, as the degree of extra-local control increases, the ambiguity of the information is stressed and a narrow range of evaluation takes place (Proposition 7). Finally, as the ambiguity and narrowness increase and as the information use is more external, the status of the unit or personnel responsible for evaluation decreases (Proposition 8).

THE TEST

Since the analysis is preliminary at this point, full discussion of the results awaits additional work with the full data set. For now, a brief discussion of the CSE sample, the variables created for this analysis, and a very preliminary presentation of regression results serve to provide the framework for further study.

The Sample

There are two basic survey samples, as described earlier in the monograph's Introduction. First there was a postcard survey concerning presence or absence of evaluation units in a nationwide sample of school districts. Second, a detailed survey instrument was sent to the heads of evaluation units in all districts having 10,000 pupils, and to a 50 percent sample of districts having 5,000 to 9,999 students. While the 263 respondent districts varied widely on many of the characteristics measured by the survey instrument,

there is little information on the characteristics of those districts which do not have evaluation units; Table 1 exhausts the information available.

Variable Definition

There are several sets of variables crucial to this analysis. First, organizational characteristics are used largaly as given in the original data: (1) The staff and budget of the evaluation unit (EU Staff; EU Budget), and (2) the Budget and assessed value of the entire district (Dist. Budget; Dist. Value). Second, the extent of local/external resources is defined in terms of the percent white (highly negatively correlated, in turn, with free lunch provision and low student performance, both also good predictors of the inflow of federal and state funds). Third, internal role of the evaluation unit is assessed by a simple summing of the number of school policies the evaluation unit is said to effect (Table 2 presents these marginals). Fourth, tight coupling is defined in terms of the district having some general guideline affecting instructional activities. Fifth, the range of activities the evaluation unit engages in is defined as a simple sum (EU Act Range).

The largest class of variables created are those which measure internal/external emphasis in terms of time the evaluation unit devotes to each and in terms of the use of evaluation information. Since construction of these measures is complex, fuller discussion is deferred until the analysis is more complete (some of the measures will be reconstructed to better reflect the underlying variance). Basically, each measure is constructed so as to either weight the relative use (Test Use: Ext is a ratio of external to internal use of both criterion- and norm-referenced tests) or sum across a number of categories of internal/external (e.g., External Use).

Results

Evidence for the role of external funding/control in the creation of evaluation units can be found in the distribution of evaluation units across states (Lyon, 1978). While in most states roughly a quarter to a third of the school districts have evaluation units, in states which have requirements for local evaluation and/or extensive state funding of schools, the percent having evaluation units is significantly higher (see Table 1). While the size of the district is important, with larger districts more likely to have both exten-

sive external funding (especially in metropolitan districts) and evaluation units, the variance across states is at least as striking. While most states have no small (5,000 to 9,999 students) districts which have evaluation units, states with extensive funding or evaluation requirements typically have evaluation units in a quarter to a third of their small districts.

Table 1: Extent of State Regulation and Funding of Local Public Schools and Presence of Evaluation Units (N=1321)*

Regulation/Funding	# Districts with EU	Percent	N
High State Involvement	226	45.0	502
Moderate State Involvement	136	25.7	529
Low State Involvement	47	16.2	290

* Universe of 750 districts with enrollments of 10,000 or more; 50% sample of districts with 5,000 to 9,999 students (573). Response rate of 100% for larger districts (n=750); 81% for smaller districts (n=464). Table total n of 1321 reflects telephone follow up results.

Therefore, while all schools exist in an institutional environment (Meyer & Rowan, 1975), the degree of institutionalization does vary as a function of the differences in the amount of federal funds and regulations (depending on size, location, and minority/poor enrollment) and as a function of state differences in funding and control over local public schools. As Propositions 4 and 6 above state, under these conditions, as institutionalization increases evaluation units will be more likely to be established, since they constitute archetypical examples of formal structure which signals compliance to external demands for accountability.

To the extent that evaluation units serve as signals of compliance to external directives and interests, their role in internal functioning of the organization is predicted to be minimal (Proposition 5). They are loosely coupled to the rest of the organization, but not by accident (Weick, 1976); they are established to serve as a boundary unit of the organization, to negotiate a complex institutional environment, not to guide the internal task performance. There is, of course, a complex relationship with the extent of federal/state interpenetration of the local school, but as Table 2 demonstrates, it is fair to characterize most evaluation units as having

minimal internal roles. Only in the area of in-service training do over half of the evaluation units report having an internal organizational role.

Table 2: Internal Role of Evaluation Units (N=227)

Activity	Percent	N
Staff In-service Training	66.5	151
Selection of Curriculum/ Program Materials	42.3	96
Allocation of Funds	36.6	83
Facilities Planning	33.9	77
Collective Bargaining	14.5	33
Teacher Performance Review	9.7	22

There are frequently differences between the amount of effort an evaluation staff may devote to generating information and the amount of and locus of its use. For example, evaluation units devote more of their time to generating information intended for groups internal to the organization and less of their time to generating information intended for external groups (federal or state agencies; parents). However, as predicted (see Proposition 6), actual use of the information generated is roughly equal externally and internally, depending on the particular measure.

As external use of the evaluation information increases in importance relative to internal use, the unit becomes increasingly loosely coupled to the rest of the organization, with organizational characteristics becoming less important in predicting how effective the evaluation unit perceives itself to be in affecting use of its results. Following Proposition 7, the use of evaluation research becomes more problematic, both externally and internally. Inconsistent, ambiguous relations between different measures of use become apparent, and normal measures of performance of the evaluation unit, such as the range of activities it performs, become decoupled from the use of the information the evaluation unit generates.

Table 3 presents the correlation matrix of all variables used, as well as means and standard deviations. The first main results are presented in Table 4. This table is strongly supportive of the third hypothesis stated above. The higher the proportion of local funds,

Table 3: Correlation Matrix

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Local/External Resources	−.007	−.087	.079	−.236	.266	.030	.160	−.087	−.025	−.045	−.373	−.335	−.268	.110	.171
2. Inrole	x	.177	.162	.163	−.202	−.038	.084	.365	.053	−.014	.068	.066	−.076	.063	.153
3. External Use	x	x	.346	.262	−.292	−.115	−.110	.179	−.016	.061	.097	.138	.028	.086	.102
4. Internal Use	x	x	x	−.055	.068	−.065	.134	.053	.124	−.125	.054	.090	−.116	.152	.058
5. EU Time: Ext	x	x	x	x	.867	.011	−.205	.136	.105	−.032	.085	.085	.124	−.178	−.032
6. EU Time: Int	x	x	x	x	x	.031	.207	−.177	−.106	.056	−.210	−.154	−.146	.135	.017
7. Test Use: Ext	x	x	x	x	x	x	.018	.021	.038	.030	.054	.070	.023	−.061	−.101
8. EU Eval Fed	x	x	x	x	x	x	x	−.138	−.089	.099	−.202	−.116	−.109	−.011	−.018
9. EU Act Range	x	x	x	x	x	x	x	x	−.066	.096	.300	.267	.066	.126	.230
10. "Tight" Coupling	x	x	x	x	x	x	x	x	x	.030	.056	.051	.142	−.089	−.009
11. Consult Budget	x	x	x	x	x	x	x	x	x	x	.111	.109	.224	.017	.139
12. EU Staff	x	x	x	x	x	x	x	x	x	x	x	.848	.466	.177	−.089
13. EU Budget											x	x	.389	.131	−.052
14. Dist Budget												x	x	.055	−.021
15. Dist Value													x	x	x
Mean	2.01	5.29	9.65	9.42	23.23	1.47	2.03	6.72	1.13	3.08	4.77	206.18	79.33	282.28	1.41
S.D.	1.40	.88	.70	6.23	6.67	.75	.82	2.07	.74	7.90	9.94	439.72	219.25	242.54	.60
#	263	221	250	256	256	200	259	263	245	241	262	247	236	186	244

the less the time spent on evaluation for external sources and the less the external use of evaluation results. Time spent on internal evaluation and internal use is also negatively related, but not as strikingly. The evaluation unit staff is significantly smaller when local funds are a high proportion, though overall budget is not strongly related (evaluation unit or district). Finally, as might be expected, district assessed value is positively related to extent of local support.

Table 4: Regression of External and Internal Emphasis/Use of Evaluation Results and Organizational Characteristics on the Extent of Local/External Resources (R = .21)

	Unstandardized Regression Coefficients	(Standard Errors)	Standardized Regression Coefficients	F-ratios
External Use	−1.17	(3.63)	−.04	.10
Internal Use	−.55	(4.62)	−.01	.01
EU Time: Ext	−2.13	(1.46)	−.46	2.14
EU Time: Int	−1.61	(1.42)	−.37	1.30
EU Staff	−.81	(.55)	−.41	2.20
EU Budget	−.01	(.01)	−.06	.09
Dist. Budget	−.01	(.05)	−.02	.01
Dist. Value	.01	(.01)	.14	1.58
(Constant 1.39)				

Table 5 presents some strong results on the internal role of the evaluation unit. As Table 2 above shows, most evaluation units play minimal roles in formulating internal school policy, even on issues directly related to the evaluation function. What is striking in Table 5 is the strong relationship between a performance measure, EU Act Range, and internal role. While not as strong, the pattern of external/internal time and use is equally striking. Internal role is positively related to external use/time but negatively related to internal use/time, as the second hypothesis above predicted. As might be expected, tight coupling in the district is directly related to an internal role for the evaluation unit. Finally, organizational characteristics are not strongly related to the presence of an internal role.

Table 6 presents less consistent results. Rather than attempt an *ad hoc* interpretation at this point, the strong relationship between the amount of evaluation unit time spent for external purposes and

Table 5: Regression of External and Internal Emphasis/Use of Evaluation Results and Organizational Characteristics on the External Use of Evaluation Results (R = .32)

	Unstandardized Regression Coefficients	(Standard Errors)	Standardized Regression Coefficients	F-ratios
External Use	.14	(.17)	.08	.57
Internal Use	−.13	(.23)	−.06	.31
EU Time: Ext	.01	(.07)	.02	.01
EU Time: Int	−.05	(.07)	−.24	.66
Test Use: Ext	−.09	(.22)	−.04	.19
Local	.13	(.22)	.06	.34
EU Act Range	.24	(.08)	.37	9.48
"Tight" Couple	.12	(.17)	.07	.53
EU Staff	−.01	(.02)	−.12	.33
EU Budget	.00	(.00)	.06	.08
(Constant 1.82)				

Table 6: Regression of External and Internal Emphasis/Use of Evaluation Results and Organizational Characteristics on the External Use of Evaluation Results (R = .26)

	Unstandardized Regression Coefficients	(Standard Errors)	Standardized Regression Coefficients	F-ratios
Internal Use	.39	(.15)	.30	7.10
EU Time: Ext	.05	(.05)	.36	1.39
EU Time: Int	.01	(.04)	.07	.04
Test Use: Ext	−.20	(.14)	−.15	1.91
EU Eval Fed	−.18	(.12)	−.18	2.29
EU Act Range	.01	(.05)	.02	.02
Local/Ext Res.	.00	(.002	.02	.03
"Tight" Couple	−.08	(.11)	−.08	.55
Consult Budget	.01	(.01)	.09	.64
EU Staff	−.01	(.02)	−.10	.16
EU Budget	.00	(.00)	.20	.86
(Constant 1.37)				

the external use of evaluation results provides some independent validation of the variables created. However, the strong negative relationship between the internal evaluation unit's evaluation of

federal programs (not outside consultants or other school units) and external use is problematic. Further investigation of the external use measure seems appropriate, even though this table presents results which are largely supportive of the argument made here.

References

Billings, R. S., Klimoski, R. J., & Breaugh, J. A. The impact of a change in technology on job characteristics: A quasi-experiment. *Administrative Science Quarterly*, 1977, 22(2), 318–339.

Callahan, R. E. *Education and the cult of efficiency.* Chicago: University of Chicago Press, 1962.

Cohen, M. D., March, J. G., & Olsen, J. P. A garbage can model of organizational choice. *Administrative Science Quarterly*, 1972, *17*(1), 1–25.

Comstock, D. E., & Scott, W. R. Technology and the structure of subunits: Distinguishing individual and workgroup effects. *Administrative Science Quarterly*, 1977, 22(2), 177–202.

David, J. L. *Local uses of Title I evaluations.* Research Report EPRC 21. Educational Policy Research Center. Stanford Research Institute, 1978.

Dornbusch, S. M., & Scott, W. R. *Evaluation and the exercise of authority.* San Francisco: Jossey-Bass, 1975.

Dowling, J., & Pfeffer, J. Organizational legitimacy. *Pacific Sociological Review*, 1975, *18*(2), 122–136.

Gouldner, A. W. *Patterns of industrial bureaucracy.* New York: Free Press, 1954.

Hasenfeld, Y. People processing organizations: An exchange approach. *American Sociological Review*, 1972, 37, 256–263.

Kaufman, H. *The forest ranger: A study in administrative behavior.* Washington, D.C.: Johns Hopkins Press, 1960.

Kirst, M. *The politics of education at the local, state, and federal levels.* Berkeley, California: McCutchan, 1970.

Leifer, R., & Huber, G. P. Relations among perceived environmental uncertainty, organizational structure, and boundary-spanning behavior. *Administrative Science Quarterly*, 1977, 22(2), 235–247.

Lyon, C. LEA program evaluation units: Preliminary demographics. Internal memo, Center for the Study of Evaluation, University of California, Los Angeles, 1978.

Lyon, C., Doscher, L., McGranahan, P., & Williams, R. *Evaluation and school districts.* Center for the Study of Evaluation, University of California, Los Angeles, 1978.

Meyer, J. W. *Research on school and district organization.* Paper presented at the Sociology of Education Conference, San Diego, April, 1977.

Meyer, J. W. The impact of the centralization of educational funding and control on state and local organizational governance. Paper presented at the HEW School Finance Study Meeting on Resource Allocation, Service Delivery, and School Effectiveness, September, 1979.

Meyer, J. W., & Rowan, B. *Notes on the structure of educational organizations.* Paper presented at the annual meeting of the American Sociological Association, San Francisco, 1975.

Meyer, J. W., & Rowan, B. Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 1977, *83*, 340–363.

Meyer, J. W., & Rowan, B. The structure of educational organizations. In M. Meyer, et al (Eds.), *Environments and organizations*. San Francisco: Jossey-Bass, 1978.

Meyer, M. Organizational structure as signalling. *Pacific Sociological Review*, 1980, *22*(4), 481–500.

Mohr, L. B. Organizational technology and organizational structure. *Administrative Science Quarterly*, 1971, *16*(3), 444–457.

Pfeffer, J. Merger as a response to organizational independence. *Administrative Science Quarterly*, 1972, *17*, 382–394.

Pfeffer, J., & Salancik, G. *The external control of organizations: A resource dependence perspective*. New York: Harper, 1978.

Scott, R. W. Organizational structure. *Annual Review of Sociology*, 1975, *1*, 1–20.

Spence, A. M. *Market signaling: Informational transfer in hiring and related screening processes*. Cambridge, Mass.: Harvard University Press, 1974.

Thompson, J. D. *Organizations in action: Social science basis of administrative theory*. New York: McGraw-Hill, 1967.

Tolbert, P. S., & Zucker, L. G. *Institutional sources of formal structure: Diffusion of civil service reform*. University of California, Los Angeles, Unpublished manuscript, 1980.

Tyack, D. B. *The one best system*. Cambridge, Mass.: Harvard University, 1974.

Weick, K. E. Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 1976, *21*, 1–19.

Wildavsky, A. The self-evaluating organization. *Public Administration Review*, 1972, *32*(5), 509–520.

Wise, A. E. Why educational policies often fail: The hyper-rationalization hypothesis. *Curriculum Studies*, 1977a, *9*, 43–57.

Wise, A. E. The hyper-rationalization of American education. *Education*, 1977b, *8*, 2–6.

Woodward, J. *Industrial organization: Theory and practice*. London: Oxford University Press, 1965.

Zucker, L. G. The role of institutionalization in cultural persistence. *American Sociological Review*, 1977, *42*(5), 726–743.

Zucker, L. G. *Interorganizational relations and transmission of cultural change: Organizational definition of local policy environments*. Unpublished manuscript, Los Angeles, 1979.

Zucker, L. G. The impact of Proposition 13 on public funding and services for education and health in California. Berkeley, California: California Policy Seminar, 1980.

Notes on Grusky's "Role Conflict and Ambiguity in Evaluation"

Introduction

It has long been recognized that environmental factors affect organizational functioning, but within the past decade organizational researchers have rapidly increased their understanding of the elements and dynamics of this process. In his chapter, Oscar Grusky, professor of sociology at UCLA, notes that "theorists in the fields of contingency, resource dependence, ecology, political economy, and open systems . . . all emphasize in one form or another the impact of the environment on the focal organization." He cites Hall's *Organizations: Structure and Process* (1977), Pfeffer and Salancik's *The External Control of Organizations* (1978), Aldrich's *Organizations and Environments* (1979), Zald's *Power in Organizations* (1969), and Katz and Kahn's *Social Psychology of Organizations* (1966) as examples. Before presenting his re-analysis of the data collected in CSE's survey of school district evaluation units, Grusky discusses role conflict and role ambiguity as aspects of organizational functioning.

Argument

Grusky's interest is in both the school district as a parent organization and in the evaluation unit as a sub-organization within the district. He is particularly concerned about the effects of the parent organization and of the sub-organization on individual role conflict and role ambiguity as perceived by the directors of evaluation units.

He characterizes school district evaluation units as, typically, new within their parent organizations and possessing limited resources. He describes the parent organizations' demands on these units for information as being somewhat beyond the units' capacity for adequate response. He notes that the unit directors' roles can be characterized as boundary-spanning; that is, cutting

across internal lines of demarcation. Since research on "boundary spanners" in other organizational settings provides reasons to expect that such individuals will experience both role conflict and role ambiguity, Grusky uses the CSE survey data to explore the effect of district and evaluation unit context on these two dimensions.

Grusky defines "role conflict," following Rizzo, House, and Lirtzman (1970), as occurring when the occupant of a position encounters inconsistent demands and expectations from his or her environment. Grusky adapted six items from a Rizzo, House, and Lirtzman measure and had them included in the CSE survey. He found that on three items, more than half of the evaluation unit directors selected the high role conflict responses agreeing that "I have received assignments without the manpower to complete them"; "I have to do things that should be done differently"; and "I receive assignments without adequate resources to execute them."

Grusky defines "role ambiguity," following Rizzo, House, and Lirtzman (1970), as occurring when the occupant of a position lacks appropriate role-related information as to duties, activity, and allocation of time and goals. Most unit directors answered survey items related to this issue in ways to suggest that they felt little role ambiguity.

Grusky suggests that negative correlation between role conflict and ambiguity may be due to the amount of work done by the evaluation units: "The harder one must work . . . the less uncertain one is (or has time to be) about one's job."

Grusky's analysis of the survey data shows that there is considerable variation in expressed role conflict and role ambiguity among evaluation directors. He explains these variations in terms of three aspects of the district context and three aspects of the evaluation unit's context. For the district, he looks at the variables of formalization, size, and heterogeneity. For the evaluation units the variables are history—the length of time the unit had existed, the percent of district funds allocated to the unit, and the number of staff in the unit. He finds that the three district variables—size, formalization, and heterogeneity—were significantly correlated with both ambiguity and conflict for individuals. For instance, the larger the district, the greater the individual's perception of role conflict. Grusky suggests that size may be related to organizational differentiation which in turn leads to lack of consensus and, therefore, role conflict. For the unit variables he finds a slight negative correlation between history and ambiguity; that is, the longer the

unit had been in existence, the less ambiguity was experienced by the director. Grusky finds no significant relationship between the unit variables and role conflict.

In his re-analysis of the CSE data, Grusky is also interested in the extent to which a relationship existed between perceptions of role conflict and ambiguity among unit directors and those who were reported as users of evaluation unit data. When the evaluation director reported both superintendents and principals as data users, he or she also reported experiencing little role conflict. This may indicate that these evaluation directors saw themselves as performing the jobs expected of them by the clients most important to them. However, when the evaluation head reported a wide range of consistent users, he or she also reported some role ambiguity. The ambiguity may have resulted from the diversity of clients served or from the skepticism of some of these clients toward data from evaluation and testing.

Contribution and Importance

The position of director of an evaluation unit is not an easy one. Such individuals appear to experience high levels of role conflict. Grusky interprets the concomitant finding that they also experience low levels of role ambiguity as meaning that they have little power within the district.

Grusky's chapter points up important although hitherto largely neglected areas of inquiry for the researcher interested in evaluation utilization. The peripheral position of the evaluation unit within the parent organization, as well as the inconsistent demands and expectations placed on the directors of such units, suggest that evaluative activities, no matter what their intrinsic merit or worth, may often be ineffectual. If school districts, in fact, wish to increase the use of testing and evaluation data internally, then they must become aware of the intra-organizational dynamics contributing to role conflict and role ambiguity. Although some factors, such as size and heterogeneity of district, may not be directly controllable by district management, others, such as resources or clear specification of evaluation unit clientele, may be.

Role Conflict and Ambiguity Among School District Evaluation Unit Heads[1]

Oscar Grusky
University of California, Los Angeles

This chapter seeks to explain variation in conflict and ambiguity among the national sample of directors of school district research and evaluation units which formed the basis of CSE's study of these units (Lyon, Doscher, McGranahan, & Williams, 1978). The chapter argues that variation in evaluation unit directors' role conflict and ambiguity is a function of school district *and* evaluation unit characteristics since both sets of organizational features influence the social context within which the director functions.

School district research and evaluation unit heads face complex administrative problems. Their organizations are, typically, both new and small (Lyon, Doscher, McGranahan, & Williams, 1978). The resources they have available, in part because of the unit's newness and size (which makes competition with other units difficult), are scarce. At the same time, the demands placed upon them by powerful persons and organizations in their school district and elsewhere are extensive and growing. Federal, state, county, and other units increasingly require information from school districts concerning the effectiveness of program functioning in specified areas.[2] Moreover, since the field of evaluation research itself is somewhat recent, unit directors' background and training are frequently in other areas. This combination—inadequate resources, increasing service demand, and minimal job training—provides all the ingredients necessary for deep-seated role conflict and ambiguity.

[1] I am grateful for the assistance of Catherine Lyon and for the support, financial and otherwise, of the CSE staff, particularly Eva Baker and Adrianne Bank. This project was also partially supported by NIMH (MN-14583). I am most appreciative of the research assistance of Pamela Tolbert and the typing of Andrea Anzalone. John Meyer and Mary Ann Millsap provided many helpful comments on the manuscript.

[2] For example, Stufflebeam, et al. (1971) write: "As a response to *outside* pressures, many school districts have installed or are now installing evaluation units" (p. 268, emphasis added).

Increasingly, organizational theorists are recognizing the significance of environmental factors on organizations. Perrow (1979) has referred to this emphasis as a "new wave-gathering force." Theorists in the fields of contingency, resource dependence, ecology, political economy, and open systems (Hall, 1977; Pfeffer & Salancik, 1978; Aldrich, 1979; Zald, 1969; Katz & Kahn, 1966) all emphasize in one form or another the impact of the environment on the focal organization. This emphasis is particularly important to the understanding of leader behavior in newly-created boundary-spanning organizations (Aldrich & Herker, 1977) such as research and evaluation units. The research evidence suggests that boundary-spanners experience high levels of role conflict (Organ, 1971, 1976; Organ & Green, 1972; Adams, 1976; Miles, 1976). The evidence is less conclusive regarding the relationship between boundary-spanning and role ambiguity. The most systematic study of ambiguity is that of March and Olsen (1976). March and his colleagues not only view ambiguity as closely linked to the choice process but, in addition, assert that it is endemic to public and educational organizations (Cohen & March, 1974). In a recent review of role conflict research Whetten (1978) observed that ". . . what is significant about the literature on boundary spanning is the noticeable lack of interest in systematically exploring the sources of role conflict." With the exception of the March and Olsen study, the same could be said about research on role ambiguity.[3]

APPROACH AND HYPOTHESES

Although role conflict and ambiguity are related (see Kahn et al., 1964), they are not identical. Conflict arises from the quality and quantity of demands placed on persons while ambiguity refers simply to perceived uncertainty. The relationship between these variables is largely unexplored. March and Olsen (1976) claim that "individuals find themselves in a more complex, less stable, and less understood world than that described by standard theories of organizational choice; they are placed in a world over which they often have only modest control" (p. 21). Ambiguity in educational

[3] March and Olsen (1976) refer to four types of ambiguity: the ambiguity of attention, ambiguity of understanding, ambiguity of history, and the ambiguity of organization. The ambiguity measure we used doesn't begin to do justice to the richness of this typology.

organizations, especially among leaders in this type of social system, is the name of the game. As for role conflict, the pioneer empirical study of the phenomenon by Gross, Mason, and McEachern (1958) was of school principals. Given the newness and instability of school evaluation units both role conflict and ambiguity should be extant in our sample.

Contrary to the standard portrait of schools as unsuccessful organizations, Meyer (1977) pictures them as highly successful because they have survived and even substantially expanded their resource base. The basis for their success is their conformity to society's institutionalized rules and the fact that they have become "relatively decoupled from the technical work of instruction." Unlike business firms which carefully control their technical structures, schools leave their instructional tasks relatively unevaluated and uncontrolled. Since evaluation units may be technically responsible both for student testing and for reviewing instructional programs, we might expect that evaluation unit directors face far more conflict and ambiguity than their counterparts in business firms. Their work has actual or potential relevance for the school's most fundamental tasks.

Our approach emphasizes the preeminent effect of context or structure on the organizational subunit and thence on role conflict and ambiguity. Three school district variables influence the context within which evaluation units function: formalization, size, and heterogeneity. Each of the variables affects the extent of interest group pressures likely to be experienced by the unit director.

"Formalization" refers to rules; its opposite is anomie. The more regulated the district organization the greater the unit's administrative control over uncertainty. The more formalized the organization, the more protected the unit head feels (Gouldner, 1954; Miles & Perreault, 1976) and the less likely he or she experiences role conflict. Both district size and heterogeneity influence diversity of interest groups in the district.

Kahn et al. (1964) note that persons in positions that link units are more likely to be subjected to conflicting requirements and pressures because they interact with persons who have competing goals and standards. However, formalization should relieve some of this conflict insofar as rules closely specify task and goal responsibilities. A district's size and heterogeneity affect its political capabilities and its ability to capture resources from the society. While, on one hand, large size and heterogeneity demand respect and hence enable districts to command greater amounts of

resources, on the other hand, they imply more competing interest groups; the greater the diversity of interest groups in a school district, the more likely there will be conflict among them. Hypothetically, interest group conflict in the school district should lead to role conflict and ambiguity among evaluation unit heads.

The unit variables of concern are history and resource availability. Pfeffer and Salancik (1978) suggest that survival is the ultimate test of organizational effectiveness and history is inextricably linked to that concept. In the present study, history was measured by length of time the unit has existed; resource availability was measured specifically by budget and number of staff personnel. History and unit resources are substantially determined by decisions external to the focal unit. Pfeffer also proposed that persons have less effect on organizations than does the institutional context because selection processes ensure homogeneity among leaders. Leaders are seen as having little discretion, anyway; since the major impact on outcomes stems from resource availability and, in school districts in particular, this is generally outside the unit head's control (Leiberson & O'Connor, 1972). Since a unit director's uncertainty mainly revolves around resources, we might have anticipated that unit variables would have a greater impact on ambiguity than on conflict.

Conflict

"Roles" are generally defined as sets of expectations about behavior associated with organizational positions. Role conflict takes place when the occupant of a position encounters inconsistent demands and expectations. Four types of role conflict have been identified by Rizzo, House, and Lirtzman (1970):

> 1. Conflict between the focal person's internal standards or values and the defined role behavior . . . 2. Conflict between the time, resources, or capabilities of the focal person and defined role behavior . . . 3. Conflict between several roles for the same person which require different incompatible behaviors . . . 4. Conflicting expectations and organizational demands in the form of incompatible policies, conflicting requests from others, and incompatible standards of evaluation.

Rizzo, House, and Lirtzman developed the factorially identifiable and independent measures of role conflict and ambiguity that we adapted for use in the present project. Six items with the highest

factor loadings were selected from their larger set. The items, listed with percent agreement in our sample, were as follows (see the monograph Introduction for a description of the sample selection):

Items	% Agree or Strongly Agree
I receive assignments without the manpower to complete them.	65
I work under incompatible policies and guidelines.	21
I have to buck a rule or policy in order to carry out an assignment.	20
I receive assignments without adequate resources and materials to execute them.	51
I have to do things that should be done differently.	53
I receive incompatible requests from two or more people.	27

It should be noted that over 50 percent of the respondents in our study selected the high role conflict response in three of the six items. Two of these three were concerned with inadequate resources.

Ambiguity

"Role ambiguity" refers to the situation that takes place when the occupant of a position lacks the appropriate role-related information. This occurs when the position is not clearly defined or when access to needed information is impeded (for example, because of the occupant's inexperience or because of the newness of the position in the organization). Specifically then, ambiguity refers to the degree of felt certainty regarding one's duties, authority, allocation of time, and goals. To measure ambiguity the five items with the highest factor loadings were selected from the Rizzo, House, and Lirtzman set. The items, listed with percent agreement in our sample, were as follows:

Items	% Disagree or Strongly Disagree
I feel certain about how much authority I have.	18
I have clear, planned goals and objectives for my job.	12
I know that I have divided my time properly.	30
I know what my responsibilities are.	10
I know exactly what is expected of me.	20

It is apparent from the above that role ambiguity was less common among directors than was conflict. A large proportion of the sample reported little ambiguity. Although we lacked comparative data, these findings would seem to contradict March and Olsen's claim regarding the pervasiveness of ambiguity. Consistent with Rizzo, House, and Lirtzman, we found a significant negative correlation between role conflict and ambiguity. ($r=-.19$, $p<.001$. See Table 1.) It may be that conflict produces expectational clarity. That is, the harder one must work and the more corners that must be cut to get the work done, the less uncertain one is (or has time to be) about one's job.

The evaluation unit heads were, typically, highly experienced and professionally trained individuals. Sixty-five percent held the doctorate, most usually in administration, elementary or secondary education, statistics, or educational or general psychology. Almost three out of ten had been school principals and over half had once been elementary or secondary school teachers. Very few (14.4%) had had any formal course work in evaluation. No significant relationship was found between taking such courses and role conflict or ambiguity.

One important function of school district evaluation and research units is to provide information of value to school administrators. Most of these units report directly to the superintendent or through one intermediary. The job involves monitoring school programs indirectly and emphasizes testing student achievement.

School District Variables

Three variables were used: formalization, size, and heterogeneity. Hage and Aiken (1970) and Hall (1977) define formalization

as the rules and procedures organizations establish to handle contingencies. All unit heads were asked to report the extent to which there were written school board policies in six areas: student conduct in classrooms, introduction of instructional innovations, type of curricular material to be used, student conduct on school grounds, instructional methods teachers use, and criteria used in evaluating student learning. This was consistent with Pugh et al.'s (1968) definition of formalization as "the extent to which rules, procedures, instructions, and communications are written." A factor analysis of the scale resulted in one factor (unnamed) that explained 39 percent of the variance of the items. Average item-item correlation was .26. Cronbach's (1951) Alpha was .68, indicating replicability and reliability.[4]

School districts were classified by size into four groups: metropolitan districts (enrollment, 45,000 or more); large districts (enrollment, 25,000–44,999); medium districts (enrollment, 10,000–24,999); and small districts (enrollment, 5,000–9,999). Existence of an evaluation unit was positively related to size. (Districts under 5,000 students were excluded from the study.)

"Heterogeneity" referred mainly to the ethnic-racial student mix in the district. The measure selected was percent of students eligible for the nationwide free-lunch program. As Table 1 shows, this measure correlated significantly with percent White, percent Black, percent Hispanic, and percent students scoring in the bottom quartile.

Table 1. Correlations Among Indicators of District Heterogeneity

	Percent Black	Percent Hispanic	Bottom Quartile Students	Percent Free Lunch
Percent White	−.77**	−.47**	−.63**	−.76**
Percent Black		−.12**	.56**	.69**
Percent Hispanic			.27**	.25**
Bottom Quartile Students				.66**

* $p < .05$
** $p < .001$
Tests of significance are two-tailed.

[4]The correlation matrix and factor loadings may be obtained from the author upon request.

Blau (1977) defines heterogeneity as "the distribution of a population among groups in terms of a nominal parameter" (p. 9). He lists thirteen nominal parameters: sex, race, religion, ethnic affiliation, clan, occupation, place of work, place of residence, industry, marital status, political affiliation, national origin, and language. The greater the number of groups and the more evenly a population is divided among them, the greater the heterogeneity. The free-lunch program is based on willingness to participate. Using this measure as a heterogeneity index probably maximized the ethnic-racial mix as well as the mix on marital status, national origin, and language. On the other hand, it may well be associated with economic homogeneity. Unfortunately, data were unavailable to ascertain the association of the index with each of these variables.

Evaluation Unit Variables

"History" referred to the length of time the unit was in existence. As anticipated, most were new organizations. Over one-third (35%) were five years old or less while 62 percent were ten years old or under. Only about one-seventh of the units (14%) had been in existence fifteen years or longer.

Two indexes of resources were used. Monetary resources were measured by the unit's allocated percentage of the school district's yearly budget. In general, the larger the unit's percent of the budget the greater the amount of slack resources (defined by Cyert & March, 1963, as the difference between existing resources and activated demands.). Personnel resources were determined by the number of full-time staff in the unit. In 23 units only part-time staff were employed; in 108 there was only one full-time employee; and 81 units ranged in size from two to five full-time persons. The largest unit reported 90 staff members. It may be assumed that the larger the staff the greater the personnel resources and the more slack.

RESULTS

In Table 2 the means, standard deviations, N's, and intercorrelations are presented for the principal variables used in the study. The findings show that the three district variables—size, formalization, and heterogeneity—were significantly related both to ambiguity and conflict, while the unit variables—history, budget, and staff—correlated with ambiguity but not conflict.

Table 2. Means, Standard Deviations, N's, and Correlations of Dependent and Independent Variables

Variable	N	Mean	Standard Deviation	Ambiguity	History	EU Budget	EU Staff	District Size	District Formalization	District Heterogeneity
Conflict	263	2.23	.69	−.19***	−.004	−.03	.07	.16***	−.08*	.12**
Ambiguity	263	2.96	.47	—	−.07*	.19***	.16**	.03	.10**	.13**
History	249	68.33	10.36	—	—	.16**	−.10**	−.16**	−.01	−.08*
EU Budget	223	33.83	42.68	—	—	—	.31***	.007	−.04	.06
EU Staff	262	4.77	9.94	—	—	—	—	.44***	−.03	.25***
District Size	263	2.51	.99	—	—	—	—	—	.002	.33***
District Formalization	263	1.95	.42	—	—	—	—	—	—	−.16**
District Heterogeneity	225	24.74	19.79	—	—	—	—	—	—	—

*p < .10
**p < .05
***p < .001
Tests of significance are 2-tailed

For the evaluation unit variables, a slight negative correlation was found between history and ambiguity. The longer the unit had been in existence, the less ambiguity was experienced by the director. A modest relationship was found between the two resource indexes and ambiguity. Budget and staff availability, which not surprisingly were positively correlated with one another, generate increased demands on the director. In March and Olsen's (1976) terms, slack provides solutions for problems and sufficient participants for each and every choice. The greater the amount of slack resources, the more problems for the director and hence the greater his/her uncertainty as to how to resolve them.

No significant relationships were found between the three unit variables and role conflict. Two opposing hypotheses were possible: that new units would produce more role conflict than old ones in that the former, being less institutionalized, would be less able to reconcile incompatible demands and pressures; or, alternatively, that old units would experience greater role conflict since they had had more time to become known, would thereby generate more demands from external units and, hence, experience greater pressures than new units. However, neither history nor slack engendered inconsistent demands and expectations for the director. It might have been anticipated that slack would increase the director's role conflict since demands increase when more resources are available, that insofar as these demands outrun resources, conflict results. This model suggests that the relationship between slack and conflict is curvilinear.

The bivariate relationships between school district characteristics and the dependent variables were, with one exception, statistically significant. District size was related to role conflict but not to ambiguity. Kahn et al. (1964) also found a significant correlation between size and role conflict. Size has been related to structural elaboration (Meyer, 1972) and to subgoal development (Dearborn & Simon, 1958), both of which are indexes of differentiation. Differentiation creates a lack of consensus which generates role conflict for the administrator.

Formalization was negatively related to role conflict which may mean that rules act as intended in regulating expectations and enhancing consensus. However, formalization was positively related to ambiguity. That is, the more rules, the greater the director's uncertainty—obviously not the intended function of rules. It may be that large numbers of rules and policies are so cumbersome and complex that they induce uncertainty among heads of units.

Heterogeneity was the final context variable. Not surprisingly, it was signficantly correlated with district size. This was reassuring since large metropolitan districts should be the most diverse and small ones least diverse. Heterogeneity was positively related to role conflict. This finding supports Thompson's (1960) theory which asserted that heterogeneity of organization members generates role diversity which, in turn, causes organizational conflict. Organizations with heterogeneous populations develop numerous "latent roles" which present complex management problems. The same finding would be predicted by political economy and resource dependence theory (Pfeffer & Salancik, 1978; Zald, 1969): the greater the heterogeneity of the district, the more diverse and extensive are competing groups. The more there are of such groups, all seeking to assert their interests, the more conflict experienced by the unit head. His or her task is complicated under such circumstances as he or she seeks to reconcile demands for information from teachers, community groups, parents, program directors, the school board, principals, administrators and, in the case of desegration, the courts. Heterogeneity was also positively correlated with ambiguity. One explanation is that heterogeneity leads to increased needs for information input and for distribution of output. The greater the number of such demands the less certain the director is regarding duties, authority, time allocation, and objectives; hence, the greater the felt ambiguity.

The regression analyses were designed to tell us how much of the variance in conflict and ambiguity the complete set of independent variables explained. The regression equation used took the following basic form: Conflict = $a + b_1$ (School District Variables) + b_2 (Unit Variables) + Ambiguity + Error.

The independent variables were regressed in stepwise fashion; first on conflict and then on ambiguity. The district variables were entered first since presumably they were less controllable by the directors than were the unit variables. Tables 3 and 4 present the main findings. The multiple R for the equations ranged from .18 to .35 indicating that the independent variables accounted for only about 3 percent to 12 percent of the variance in conflict and ambiguity. Obviously, this was not a great deal. The more conservative adjusted R^2 measures, which consider the number of variables in the equation, reduced this amount further.

To summarize this section and the results thus far, we anticipated that a selected set of school district variables and evaluation unit variables would explain role conflict and ambiguity among unit

Table 3. Regression of School District and Evaluation Unit Variables on Role Ambiguity

Independent Variable	Unstandardized Regression Coefficient	Standard Error	Standardized Regression Coefficient
Heterogeneity	.306	.001	.132
Size	−.233	.038	−.051
Formalization	.158	.078	.144
History	−.308	.003	−.065
EU Budget	.155	.0008	.146
EU Staff	.454	.003	.105
Conflict	−.128	.048	−.192
(Constant)	3.067		
Multiple R	.350		
R Square	.122		
Adjusted R Square	.088		

Table 4. Regression of School District and Evaluation Unit Variables on Role Conflict

Independent Variable	Unstandardized Regression Coefficient	Standard Error	Standardized Regression Coefficient
Heterogeneity	.366	.002	.105
Size	.138	.057	.200
Formalization	−.148	.120	−.090
History	.143	.005	.020
EU Budget	.161	.001	.0001
EU Staff	.948	.005	.014
Ambiguity	−.293	.111	−.194
(Constant)	2.888		
Multiple R	.337		
R Square	.113		
Adjusted R Square	.078		

directors. In fact, we were able to explain only a small portion of the variance. We turned, next, to the problem of organizational influences on the use of evaluation information. The main question we sought to answer here was: Is there a relationship between the reported users of evaluation data and role conflict and ambiguity among unit heads?

Conflict, Ambiguity, and Use of Evaluation Data

Lyon, Doscher, McGranahan, and Williams (1978) provide extensive information from the CSE study on what evaluation unit heads do. They found that 95 percent of the unit directors ranked assessing student achievement as one of their most time-consuming activities. Seventy-four percent ranked it as the most time-consuming. Seventy-five percent of the directors claimed that testing was the major data collection method. Moreover, most unit heads reported that almost half their time, 46 percent, was spent on early childhood and elementary education. The authors of the CSE report concluded that ". . . the survey and fieldwork confirm the continuing dominance of testing in all activities of evaluation offices . . ." (1978, p. 100).

David's (1978) intensive field study of school district use of Title I evaluations found that they ". . . do not primarily serve either as a means of judging the program or as a guide to program improvement" (p. v). This was so for three reasons: (1) Most programs were stable; only minor changes took place anyway. (2) Typically, evaluation results were received too late to be useful; and (3) Other factors, such as political demands, played a key role in program change. If these findings hold up when replicated then we must assume that these units mainly meet reporting requirements and do not play a significant part in program change. This may be because there isn't a great deal of program change other than that which comes about as a result of externally-imposed legislation.

Respondents in the CSE study were asked to identify the major users of their units' reports. It was found that the consistent users were program directors (62%), superintendents (60%), central office staff (58%), and principals (52%). Only one-third reported teachers as consistent users. This was about the same percentage reporting federal and state agencies as users. It was evident that the units mainly serve the school administration.

As Table 5 demonstrates, role conflict was negatively correlated with service use by superintendents and/or principals; that is, those who did not report these parties as consistent users were most likely to experience high conflict. This suggests that the closer the service ties between the evaluation unit head and the school superintendent and/or principals, the less conflict was experienced. In other words, the way unit heads reduced stress was by accommodating to those who have administrative authority or influence.

Table 5. District Use of Evaluation Unit Data and Role Conflict and Ambiguity

Reported Consistent User	Role Conflict	Role Ambiguity
Superintendent	−.20 (p=.02)	.19 (p=.02)
Principals	−.18 (p=.03)	.12 (p=.10)
Board members	--------	.19 (p=.001)
Parents or local citizen groups	--------	--------
Teachers	--------	−.21 (p=.007)
Central office staff	--------	−.15 (p=.04)
Federal agencies	--------	.10 (p=.003)
State agencies	--------	--------
Program Director	--------	.13 (p=.07)

A different pattern was found for ambiguity. Unit heads with high ambiguity were more likely to report superintendents, principals, program directors, board members, and federal agencies as consistent users of their services and less likely to report teachers and central office personnel as users. It appears that the greater the range of perceived use of evaluation services, the more the felt ambiguity. Any type of administrative contact can generate uncertainty, but contact with those highly placed in the organization (such as superintendents, principals, board members) was associated with high ambiguity while contact with lower level roles (teachers, central office staff) was associated with low ambiguity.

Ambiguity occurs when shared role specifications are incomplete—the officeholder is unsure of what is desired or how to behave. Unit directors felt most uncertain when consistent users were principals, superintendents, board members, program directors, and federal agencies. Perhaps this was because these officials not only have organizational clout, but also have little confidence in the test data the evaluation units produce. Their profound lack of confidence in the units' major product was described and analyzed by David (1978) who quoted several officials' telling criticisms of standardized test results:[5]

> How can you evaluate when kids are starting at different places and developing at different rates? Means don't mean anything. (Director)

[5] O'Reilly amplifies these kinds of criticisms elsewhere in this monograph.

> Individual diagnostic tools provide the basis for my judgment of program success; not the standardized tests. (Principal)
>
> If the standardized test scores are negative, it's okay because everyone buys the argument that they can be discredited. (Administrator)

David (1978) also reported that teachers were critical of testing. However, unit heads' ambiguity was low when teachers were consistent users, probably because they were less threatened by teacher criticisms. Being held in low regard by their superiors understandably carried more weight.

CONCLUSIONS

Juvenal wrote in his *Satires*, "But who is to guard the guardians themselves?" while Plato, much less the realist, stated in the *Republic*, "What an absurd idea—a guardian to need a guardian." Evaluation is a booming enterprise and evaluation units in school districts are to be found in many districts of substantial size. These units are conceived by some to be public guardians, data collectors, and assessors.

This chapter argues that the social resource characteristics of the school district—that is, the external context within which evaluation units function, and the organization of the unit itself—are key sources of information about them and particularly about the amount of conflict and ambiguity confronted by the directors. Contrary to expectations, school district and evaluation unit variables did not explain much of the variance in the directors' role conflict and ambiguity. However, our findings do suggest that evaluation unit heads fill a very difficult position in the school district and that a key source of their difficulties is that their main output is not highly regarded by their superiors. Still another problem stems from the fact that they have limited contact with the programs they evaluate. These two problems are related because if they had better contact with major school programs they would have access to information which could enhance their organizational position and power and improve the quality of their contribution. The directors' overall level of role conflict was high while their level of role ambiguity was low. Since power and ambiguity are highly correlated, this latter finding reflects their low power. It may be that the tasks of the directors are too well established and not

ambiguous enough. If evaluation units are to make a difference in school district innovation and functioning, they must involve themselves closely in classroom activities and related programs. Although this would produce greater uncertainty for the directors it could also help make their evaluation tasks considerably more meaningful.

References

Adams, J. S. The structure and dynamics of behavior in organization boundary roles. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally, 1976. Pp. 1175–1200.

Aldrich, H. *Organizations and environments*. Englewood Cliffs, N.J.: Prentice-Hall, 1979.

Aldrich, H., & Herker, D. Boundary spanning roles and organization structure. *Academy of Management Review*, 1977, *2*, 217–230.

Blau, P. M. *Inequality and heterogeneity*. New York: The Free Press, 1977.

Cohen, M. D., & March, J. G. *Leadership and ambiguity*. New York: McGraw-Hill, 1974.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrica*, 1951, *16*, 297–334.

Cyert, R. M., & March, J. G. *A behavioral theory of the firm*. Englewood Cliffs, N.J.: Prentice-Hall, 1963.

David, J. L. *Local uses of Title I evaluations*. SRI International, Research Report EPRC 21, July, 1978.

Dearborn, D. C., & Simon, H. A. Selective perception. *Sociometry*, 1958, *21*, 140–143.

Gouldner, A. W. *Patterns of industrial bureaucracy*. New York: Free Press, 1954.

Gross, N., Mason, W. S., & McEachern, A. W. *Explorations in role analysis*. New York: John Wiley, 1958.

Hage, J., & Aiken, M. *Social change in complex organizations*. New York: Random House, 1970.

Hall, R. H. *Organizations: Structure and process* (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Kahn, R., Snoek, J. D., & Rosenthal, R. A. *Organizational stress: Studies in role conflict and ambiguity*, New York: John Wiley, 1964.

Katz, D., & Kahn, R. *Social psychology of organizations*. New York: John Wiley & Sons, Inc., 1966.

Leiberson, S., & O'Connor, J. F. Leadership and organizational performance: A study of large corporations. *American Sociological Review*, 1972, *37*, 117–130.

Lyon, C., Doscher, L., McGranahan, P., & Williams, R. *Evaluation and school districts*. Center for the Study of Evaluation, University of California, Los Angeles, 1978.

March, J. G., & Olsen, J. *Ambiguity and choice in organizations*. Bergen, Norway: Universitetsforlaget, 1976.

Meyer, J. W. *Research on school and district organizations*. Paper presented at Sociology of Education Conference, San Diego, 1977.

Meyer, M. W. *Bureaucratic structure and authority.* New York: Harper and Row, 1972.

Miles, R. H. Individual differences in a model of organizational role stress. *Journal of Business Research,* 1976, *4,* 87–102.

Miles, R. H., & Perreault, W. D., Jr. Organizational role conflict: Its antecedents and consequences. *Organizational Behavior and Human Performance,* 1976, *17,* 19–44.

Organ, D. W. Some variables affecting boundary role behavior. *Sociometry,* 1971, *34,* 524–537.

Organ, D. W. *Field studies of organization boundary roles: A selective review and notes toward some tentative prescriptions.* Paper presented at the American Psychological Association meetings, Washington, D.C., 1976.

Organ, D. W., & Green, C. N. The boundary relevance of the project manager's job: Findings and implications for R&D management. *Research and Development Management,* 1972, *3,* 7–11.

Perrow, C. *Complex organizations: A critical essay.* (Rev. ed.) Glenview, IL: Scott, Foresman, and Co., 1979.

Pfeffer, J., & Salancik, G. R. *The external control of organizations.* New York: Harper and Row, 1978.

Pugh, D. S., Hickson, D. J., Hinings, C. R., & Turner, C. Dimensions of organizational structure. *Administrative Science Quarterly,* 1968, *13*(1), 65–105.

Rizzo, J. R., House, R. J., & Lirtzman, S. I. Role conflict and ambiguity in complex organizations. *Administrative Science Quarterly,* 1970, *15,* 150–163.

Stufflebeam, D. L., et al. *Educational evaluation decision making.* Itasca, IL: F. E. Peacock Publishers, 1971.

Thompson, J. D. Organizational management of conflict. *Administrative Science Quarterly,* 1960, *4*(4), 389–409.

Whetten, D. A. Coping with incompatible expectations: An integrated view of role conflict. *Administrative Science Quarterly,* 1978, 23(June), 254–269.

Zald, M. N. *Power in organizations.* Nashville, Tenn.: Vanderbilt University Press, 1969.

Notes on O'Shea's "Evaluation Units' Problems and Possibilities"

Introduction

All organizations reflect at some times on such questions as "Are we successful in what we are doing?" "Are our techniques working properly?" "What is in need of change and improvement?" A variety of responses and methods have been developed in the private and public sectors to answer these questions, ranging from leaders' guesses or hunches, through monitoring and control procedures, through market research, and finally to elaborate management information systems.

In education, a recent response to questions of effectiveness has been the emergence of the field of educational evaluation. Over the years evaluation has developed from an offshoot of educational research to an established field with its own unique problems and procedures. A current manifestation of that growth, and an important milestone in evaluation's history, is the recently published *Standards for Evaluations of Educational Programs, Projects, and Materials* (Joint Committee on Standards for Educational Evaluation, 1981) which attempts to provide a common language and set of procedures for evaluation practitioners and theorists.

But a fundamental problem plagues educational evaluation. Evaluations use uncertain methodologies to study technically uncertain activities that take place in uncertainly understood contexts. By this we mean that evaluations, which are almost always carried out in a field setting, are necessarily limited in the kinds of research design that can be employed. Many, probably most, programs and projects that are evaluated have many factors operating at the same time that can influence, usually in indeterminable ways, the program or project's outcome.

Given these serious constraints, what is the maximum contribution, in education, that evaluation can make towards answering the age-old questions about organizational effectiveness and improvement, with which we opened these comments? It is

important for the evaluation community to address these questions so that work can be directed towards reducing the uncertainties and channelling energy into more achieveable directions.

Argument

In his chapter, O'Shea explores the built-in limits of evaluation, as it is currently practiced, as an aid to educational decision making. While O'Reilly examined decision makers and their contexts, O'Shea looks at the inherent assumptions underlying educational evaluation itself and points out that such evaluation is now frequently confined to use as a monitoring device rather than as an analytical tool, especially in a politically charged educational context.

He begins his argument with a brief but useful review of the history of those federally-funded educational programs that gave impetus to the modern development of educational evaluation. He reminds us that Sen. Robert Kennedy insisted on regarding evaluation as a *monitoring* activity which would serve to increase political support for reforming the educational system so it would more adequately serve the poor. But at that time, the advocates of Planning, Programming, and Budgeting Systems (PPBS), Management by Objectives (MBO), and related *analytical* approaches to improving decision making, as represented by Robert McNamera in the Defense Department, were in the ascendancy.

The advocacy of analytical accounting was attractive to many educational researchers who spearheaded the educational evaluation movement and who subsequently emphasized evaluation's analytical functions rather than the original monitoring role that Senator Kennedy intended.

O'Shea then discusses some conditions that must be present or assumptions that must be met if evaluation is to fulfill its potential as an analytic tool for school districts. Briefly stated, educational methodologies are technically weak—we do not know with sufficient clarity the cause and effects in educational practice. What, for example, precisely makes for effective teaching and efficient learning? The precise linkage between cause and effect is still uncertain. Given this weak technology, school districts function as institutional rather than as technical systems—they pay far closer attention to symbolic acts than to technical concerns.

Such an institutional setting, where the relationships between teaching and learning are hunches rather than immutable truths, is not an appropriate one in which evaluators should try to perform heroic, complex analyses. It follows that until more is known about the causal linkage, evaluation should realistically limit itself to achievement monitoring rather than complex analytical work.

Drawing upon data from the CSE school district study, O'Shea shows that evaluation units in the sampled districts did indeed spend the vast majority of their time in monitoring rather than in analytic activities. O'Shea suggests that if the units wish to become involved in instructional, i.e., technically oriented, evaluation they must heed Ebel's (1980) advice to adopt a methodology appropriate to that which they are evaluating, i.e., tc be more hermeneutic than experimental, more qualitative than quantitative. They must also seek out measures which are sensitive to teachers' instructional goals and which provide information useful for local instructional improvement.

Contribution and Importance

Based on the preceding kinds of argument, O'Shea encourages districts wishing to inform instructional decision making to emphasize the use of instruments which are closely linked in content to the district's instructional programs. By using instruments such as criterion-referenced tests school districts can get a more direct measure of the effectiveness of the instructional program than can be achieved with other kinds of tests. Finally, for those who wish to use evaluation for district-level decision making, O'Shea emphasizes the selection of methods, which will usually be qualitative, which are responsive to the conditions and uncertainties of the real world of schooling.

References

Ebel, R. L. *Practical problems in educational measurement.* Lexington, Mass.: D.C. Heath and Company, 1980.

Joint Committee on Standards for Educational Evaluation. *Standards for evaluations of educational programs, projects, and materials.* New York: McGraw-Hill, 1981.

School District Evaluation Units: Problems and Possibilities[1]

David W. O'Shea
University of California, Los Angeles

INTRODUCTION

Widespread emergence of school district evaluation units in recent years reflects a local response to mandatory reporting provisions of the Elementary and Secondary Education Act (ESEA), and of similar legislation in the individual states. ESEA required participating states to secure, from each school district, data on student academic achievement by selected grades. The legislation also required each district receiving federal funds

> . . . to assess the effect of its program and to identify weaknesses as well as strengths of the project, thus serving as a tool for program revision and improvement (USCAN, p. 411, cited by David, 1978, p. 4).

Eighty-five percent of school district evaluation units, which now exist in 43 percent of the nation's 750 districts enrolling 10,000 pupils or more, have been established since 1965, the first year of ESEA (see Lyon, Doscher, McGranahan, & Williams, 1978). Curiously, however, while these evaluation units expend great effort on gathering achievement data, little attention is given to the analysis of these data with a view to exploring the "weaknesses as well as strengths of the project," as legislators hoped. In practice, district evaluation units are more likely to engage in *achievement monitoring* rather than in *analytic evaluation* (see Kerins, 1973, p. 76; Ebel, 1980, pp. 281–292). Monitoring involves recording changes over time in average levels of academic achieve-

[1]I am indebted to the following persons for their thoughtful comments on earlier drafts: Adrianne Bank, Robert F. Boruch, Oscar Grusky, John W. Meyer, Mary Ann Millsap, Richard Williams, and Lynne Zucker. Also I am indebted to CSE for financial support.

ment, by selected grade levels. Analytic evaluation attempts to assess the contribution of specific instructional programs to student learning.

The predominance of monitoring activity in the work of school district evaluation units provokes two questions, both of which are addressed in this chapter. First, what inhibits analytic evaluations within school districts? Second, what might be done to facilitate more emphasis upon analytic evaluations?

In exploring factors inhibiting analytic evaluations at the school district level we begin with a brief review of the origin of ESEA's evaluation requirements, and of the attempts to implement national evaluation of Title I programs, which was the largest component within ESEA. This review indicates that a major inhibiting factor is disjuncture between the assumptions underlying most approaches to program evaluation and the principles underlying the practice of instructional programs in the schools. Evidence of such disjuncture is apparent in the content of the debates preceding implementation of ESEA in 1965, and also in the experiences of federally sponsored evaluations of instructional programs which were funded under Title I of ESEA. To explore relevant evidence, first we turn to the original debate regarding ESEA evaluation.

ASSUMPTIONS UNDERLYING ORIGIN OF EVALUATION REQUIREMENTS

Prior to passage of ESEA in 1965, most evaluations within local school districts focused upon pupils, not upon instruction. The instructional program was taken as given, leaving in question how well individual pupils were performing in their effort to master the program content. Beginning in the 1950's, the civil rights movement drew public attention to the fact that not only did individual pupils differ in their academic achievement, but whole groups among racial minorities were performing at lower levels than the average for the majority white population. Specifically, by the 1960's the phenomenon of the "achievement gap" was well documented (see especially Coleman, et al., 1966). Pupils graduating from schools in low income, predominantly minority communities, were found to have average levels of achievement three grades below the average for pupils from schools serving the more economically advantaged majority white population.

Until the findings of Coleman, et al. (1966), emphasizing the dominance of socio-economic background upon pupil achieve-

ment, civil rights leaders attributed the achievement gap to factors such as the discriminatory allocation of financial resources among schools and/or the absence of staff accountability to parents. Congress responded with legislation addressing both possibilities. To improve resources, ESEA made available over three billion dollars annually for local schools. Title I of the Act, oriented to poverty area programs, received the largest proportion of the funds allocated. To facilitate staff accountability, the Act mandated state-level collection of pupil achievement data for each school. Finally, to keep Congress apprised of the impact of the money provided, ESEA called for annual national evaluations of the Title I program.

Requirements regarding collection of pupil achievement data were written into the Act in response to demands from Senator Robert Kennedy, then representing New York (see Bailey & Mosher, 1968; McLaughlin, 1975). Robert Kennedy supported ESEA on condition that the legislation include a reporting requirement and "good faith administration efforts to hold educators responsive to their constituencies and to make educational achievement the touchstone of success in judging ESEA" (McLaughlin, 1975, p. 3). In Kennedy's opinion, if achievement data were available, parents would be in a position to monitor school performance and thus hold educators accountable for learning outcomes, creating an essentially *political* pressure for program improvement.

Kennedy's insistence upon the collection of achievement data was welcomed by officials in the Department of Health, Education, and Welfare (HEW), though not because they shared his interest in facilitating parental pressure upon the schools. It so happened that coincident with the passage of ESEA, President Lyndon Johnson, fascinated by Secretary of Defense McNamara's techniques for controlling Pentagon budgets, announced that the Defense Department's system of fiscal management, the Planning, Programming, and Budgeting System (PPBS), was to be installed throughout all sections of the executive branch (McLaughlin, 1975, p. 6).

To implement PPBS in HEW, the department which had overall responsibility for ESEA, William Gorham was brought in as Assistant Secretary for Program Evaluation (ASPE). From the outset the ASPE staff, unlike Senator Kennedy, took an analytic approach to evaluation. As Gorham (1967) wrote later (see McLaughlin, 1975, p. 7):

> Title I of the Elementary and Secondary Education Act . . . is essentially a vast experiment designed to find effective ways of reaching disadvantaged children. We know that most school systems have not been doing a very good job with these children, but there is no consensus among educators about how to do better.

Obviously, Gorham and Kennedy had quite different approaches to program assessment in mind.

> Kennedy thought of evaluation as a means to ensure that schools focused on the needs of the poor, to make sure that Title I 'worked.' Gorham's ASPE staff thought of evaluation as a way to find out 'what works,' to identify the most effective compensatory strategies (McLaughlin, 1975, p. 7).

Kennedy's approach involved achievement *monitoring*, rather than program *evaluation*. Evaluation is analytic, involving what Thompson (1967, p. 86) calls "efficiency tests," focusing upon the worth, or effectiveness, of program processes in relation to program outcomes.

Inhibiting the development of *evaluation* of instructional programs as opposed to *monitoring* schooling outcomes are contradictions between the assumptions underlying analytic evaluation and the nature of instructional processes in school settings. These contradictory assumptions become more apparent if one looks at past attempts at national evaluations of Title I programs.

NATIONAL EVALUATIONS

In enacting ESEA, Congress called for both the development of school level achievement data, and an annual evaluation of the Title I program at the national level, without specifying the form that such an evaluation might take. However, as noted earlier, the approach adopted for program evaluation was influenced by the government's desire to implement PPBS as *the* system for management of all federal programs.[2] Such an objective serves to focus evaluation upon cost-benefit analyses, orienting evaluation studies toward identification of the most cost-effective, or efficient, means of improving student learning. To facilitate cost-benefit studies it is necessary to discover which program elements actually are effective in promoting learning gains in classrooms. For this purpose the typical, though not exclusive, approach has been to base eval-

uation upon experimental design, comparing the progress of students in funded programs with the progress of others in selected control groups. As Apsler (1977, p. 14) confirms, the experimental paradigm, so fruitful in the natural sciences, is "advocated by nearly every methodological discussion of evaluation research" (see for example, Campbell, 1975; Houston, 1972; Suchman, 1967; Riecken & Boruch, 1975).

In practice, this paradigm proved to be extremely difficult to apply to national evaluations of Title I programs, a factor inhibiting downward extention of analytic evaluation to the local district level. Nationally, an early attempt at analytic evaluation was the study of compensatory education conducted for the ASPE section of HEW by the TEMPO division of General Electric. Data were obtained from schools in eleven districts considered to have "exemplary" Title I programs. It was hoped that analysis of these data would determine whether "different amounts of money spent in different ways would be significantly and differentially successful in effecting academic achievement" (McLaughlin, 1975, p. 37).

Unfortunately, the TEMPO study foundered. The main problem was that, as Ebel (1980, p. 290) concludes, in the case of schools it is almost impossible to abstract a program from its context and study it in isolation. Similarly, McLaughlin (1975, p. 37) reports:

> TEMPO analysts were able to identify neither a Title I population, nor Title I program, nor significant achievement gains that could be attributed to Title I funds. The incomplete, confusing, or non-

---

[2]The objective at the federal level to rationalize program administration did not terminate with changes in the White House. Though the move to adopt PPBS began under Johnson, officials in OE were arguing for the same objective as recently as 1977, as evidenced by the following comments of John W. Evans and Janice K. Anderson (Evans & Anderson, 1977, p. 160), who were senior officials in the Office of Planning, Budgeting, and Evaluation at the U.S. Office of Education.

> Since the late sixties, Congress has become increasingly disillusioned with the policy relevance of most general or basic research produced through the open-ended grant process, and it has demanded that federal agencies produce more immediately useful research/evaluation results. Mission-based federal agencies have responded by moving more in the applied direction, and in the process have assembled technically trained evaluation staffs to design and closely monitor the conduct of evaluations and other analyses. . . . They are part of a larger trend, with other elements being the emphasis on accountability, management by objectives, program planning and budgeting, and the creation of the new budget committee's procedures and ceilings within the Congress. The basic thrust of this larger trend is a belated effort in the social program area to rationalize federal policies, and the allocation of scarce resources.

> existent records at the local level made it difficult for TEMPO analysts to determine which services were being purchased with Title I dollars, or to specify the group receiving special attention through Title I. In practice, there seemed to be no real Title I program to evaluate.[3]

Basically, confronting the staff of TEMPO, and of subsequent Title I evaluations, was the fact that the logic of evaluation using the experimental paradigm is at odds with the logic underlying schools.[4] This same conflict of logics also frustrates attempts at the school district level to develop formal program evaluation.

---

[3] As McLaughlin (1975, p. 40) emphasizes:

> A central requirement of an impact, cost-benefit study—the ability to tie inputs to outputs—does not coincide with the operational reality of Title I. An attempt to trace the flow of Title I dollars to specific programs and outcomes is beset with problems. In what Michael Kirst calls the "Byzantine world of school accounting," it is difficult if not impossible to trace the course of Title I dollars through the school system. Some cities, especially larger cities, have over 100 sources of income. As the number of revenue sources increases, the ability of evaluators to identify the impact of any single source diminishes.

[4] In discussing McLaughlin's (1975) study of ESEA, House (1978) draws conclusions similar to those offered in this chapter. Commenting upon the approach to Title I evaluation, House (1978, p. 388) concludes that:

> The evaluation policy developed by the federal government reflected a particular ideology. It reflected the belief of systems analysts and economists that evaluation should be used primarily to detect the most efficient programs. This concern for efficiency led to evaluations that could find no differences between the new programs and those already existing in the public schools.
>
> The result has been that educational funds have been constrained at the federal level for lack of visible results and education has been discredited for its inability. Actually the lack of results reflects more the type of evaluation employed than the quality of the educational programs.

The ultimate danger of an approach to evaluation embodying a logic at odds with the reality of schooling is, as House (1978, p. 392) warns, that while evaluation is initiated in order to insure the success of educational programs, it can generate pressures directed toward the imposition on the schools of programs designed to ensure the success of evaluations. In brief, evaluation that assumes schools conform to technical rationality is likely to have the effect of absorbing the normative systems of schools into the system implicit in the experimental paradigm underlying most evaluations.

CONFLICTING LOGICS

Insight into the basis for these conflicting logics is provided by Apsler's (1977, p. 15) comments:

> Although the needs of action programs might be very specific, such as to raise children's test scores or to reduce the incidence of crime, the procedures by which these ends will be attained are generally quite *unspecific*. Unfortunately, little is known yet about how social action programs can raise children's test scores, reduce crime, and so on. As a result, social action programs often become operational without clear approaches or means of attaining their goals. Not surprisingly, the actual operations of a social action program may then be primarily determined by intuitions and exigencies. The confusing situation that develops—in which the approaches are poorly thought through and in which approaches frequently vary in response to recurrent crises and changing intuitions—presents enormous difficulties to the evaluators.

Such confusion is especially prominent, one might add, when one operates within the constraints of the experimental, or even the quasi-experimental, paradigm. Evaluation according to the experimental paradigm, with its assumptions about treatments causally related to outcomes, follows the logic of technical rationality. Schools cannot fit this mode, as educators lack a theory on the basis of which they can specify which instructional interventions are likely to produce specific learning outcomes. Were such a theory available it would guide specification of the interventions necessary to reduce the substantial gap in average levels of academic achievement between schools serving low and middle-income communities. As Averch *et al.* (1962) determined, educational research has not yet identified "what works." Instruction, like parenting, remains more an intuitive and practical art than science, as Lortie (1975) argues. Like parents, teachers follow the normatively grounded logic of *practical rationality*.

However, as Bernstein (1975, p. 64) reminds us, in our technological society we are dominated by the values of a culture that is grounded in *theoretical* knowledge, and committed to *technical rationality*. In this context, it is not surprising that most evaluators, and policy makers, subscribe to the notion that all institutions should operate in a technically rational mode, and be evaluated in terms of the experimental paradigm. However, as Goodlad (1975) argues, and also House (1978), once you begin thinking about

schools in the framework of the experimental paradigm, with its dependent and independent variables, there is a danger of absorbing the reality of schooling into the model of a technically rational system epitomized by the modern factory.[5] Factories embody technology based upon tested knowledge of means-ends relationships. In designing an automobile plant, for example, planners draw upon a body of knowledge that is sufficiently developed to serve as a basis for specifying in advance all the functions that must be performed in order to transform an array of raw materials into a finished automobile. Systems embodying technical rationality are greatly extending man's domination over nature, facilitating the exploration of space, and providing the material basis for contemporary industrial civilization.

Schools, however, are not of the same organizational genre as factories and other technically rational systems. Given the enormous prestige of the latter systems there is, of course, continuous pressure for schools to move toward the technically rational mode. Inhibiting such movement in the field of schooling is the absence of the ingredient most necessary to technically rational systems; tested knowledge of the relationships between instructional means and educational ends (see Averch et al., 1972). In effect, as Dreeben (1970) reminds us, we lack a technology of instruction, a condition leading Metz (1978, p. 20) to conclude:

> Technology, then, is a major problem for the public schools. They are faced with the task of creating changes in diverse raw material through processes that are poorly understood, in the absence of any universally effective means, and without any trustworthy way of measuring the success or failure of whatever methods they finally apply.

Schools' weak technology, coupled with their highly variable material, in the form of pupils, constrains a decentralized type of organizational structure. Coordination is achieved not by staff implementing a specific technology but by the collective enactment of a moral order, or institutionalized social system (see Metz, 1978, p. 30). As Meyer (1977) argues, schools are *institutional* rather than *technical* organizations. Coordination of activities is

[5] One then becomes entrapped in the phenomenon Wise (1977, p. 44) labels "hyperrationality." In essence, hyperrationality represents an assumption that the logic of technical rationality is at work when, in fact, it is not.

achieved primarily by basing school programs upon personnel who, having been trained, screened, and certified by state approved teacher-education programs, share a common understanding of the professional duties to be undertaken, and of the rules governing performance of these duties. Schools, as *institutional* organizations, articulate a particular set of rules, many of which are embodied in the educational codes of the different states, which are the political entities constitutionally empowered to authoritatively express societal expectations regarding the ways in which emerging generations ought to be inducted into adult society. Evaluation of the organizations within which the induction process is institutionalized, and of the specific programs within these organizations, calls for a different approach than that which is appropriate for technical organizations.

TECHNICAL VERSUS INSTITUTIONAL ORGANIZATIONS

The major determinant of organizational type, in Meyer's (1977) opinion, is the nature of the social environment within which the organization evolves. Technical organizations, such as factories, evolve in environments with complex technologies. By contrast, institutional organizations, such as schools, evolve in environments with very limited technologies but with elaborate institutional rules. The task of a technical organization is to coordinate and control technical work. The task of an institutional organization is to create structures that conform with institutionalized rules generated over time within society to order the maintenance of the social system.

Technical organizations, in their pure form, are predominantly cybernetic systems, oriented to goal attainment, and regulated by feedback from their own results (see Habermas, 1970, pp. 87–93). By contrast, institutional organizations are socialization systems, oriented to getting participants to internalize the system's constitutive norms, or institutionalized rules.

In evaluating technical organizations, priority is given to assessing the cost-effectiveness of alternative means to achieve specific outcomes. When institutional organizations are being evaluated priorities shift. The dominant concern becomes the degree to which organization structure and processes conform to the values to which the organization is expected to give expression. The effectiveness of a school in promoting the academic achievement of the pupils plays a minor role in school accreditation. The priorities are

facilities, staff qualification, teacher-pupil ratios, and the scope and content of the educational program.

Not only do technical and institutional organizations differ in terms of the premises upon which they are structured, but their structural characteristics take on radically different forms. Knowledge of means-ends relationships allows the administrators of technical organizations to direct all operations from the top down. By contrast, in the case of institutional organizations, the relative absence of technical knowledge constrains administrators to leave operations to the discretion of staff at the operational, or technical, level. One consequence in the case of schools, as Metz (1978, p. 21) concludes, is that:

> The persons who perform the actual work of the organization need to be given relatively large and diffuse tasks with the right to make important decisions independently as they use their intuition to adjust their methods to the requirements of each specific instance.

While technical organizations are highly centralized, with the administrative and technical, or production, levels tightly coupled, schools, as institutional organizations, are decentralized, with administration and the technical or teaching level loosely coupled, at least with regard to the formal work of the schools, the instruction of students (see Weick, 1976; Meyer, 1977.)[6] School principals, for example, though including instructional leadership among their responsibilities, seldom visit classrooms or involve themselves in the details of instructional processes. The latter are delegated almost entirely to teachers.

In summary, therefore, schools as institutional organizations combine a relative absence of technical rationality with a corresponding emphasis upon practical rationality, and loosely coupled relationships between administrative and instructional levels, three conditions that constrain school evaluations toward program monitoring.

---

[6] Others concerned with the nature of schools as organizations also have noted their loosely coupled structure. See, for example, Bidwell (1965); Dreeben (1970); Metz (1978); Wolcott (1977).

PROGRAM MONITORING

Given that institutional and technical organizations operate on the basis of different premises, approaches to evaluation that are appropriate for the one type prove inappropriate for the other. In exploring the theory of organizations Thompson (1967, pp. 83–101) points out that the appropriate mode of organizational assessment is determined by two variable factors: (1) the degree to which there is clarity regarding the effects, or outcomes, desired; (2) the degree of knowledge available regarding the means that actually can produce the desired outcome.

The same principles apply to program evaluation. Where it is clear just what outcome is desired, and there is full knowledge of the means needed to attain this outcome, administrators can both prescribe and control the relevant means. In this context, evaluation can focus upon the *cost-effectiveness,* or *efficiency,* of the program.

Where knowledge of means/ends, or cause/effect, relations are incomplete, as with schools and their programs,

> . . . . the efficiency test is inappropriate, for there is no way of assessing the net effect of causal action. In this case, the appropriate test is not the economic one, but the *instrumental* one—whether a desired state of affairs is achieved (Thompson, 1967, p. 86).

In the case of schools, knowledge of the causes of student academic achievement are notably incomplete, a condition that leads Ebel (1980, p. 288) to conclude:

> In seeking to evaluate an educational program, we may be led astray if we try to follow too closely the model of the scientific research study. An educational program is not a stable, natural phenomenon with built-in operating characteristics that people may discover and put to use, but not alter in any fundamental way. An educational program is a human artifact, highly complex, infinitely variable, and subject to incessant change. It defies precise definition or accurate measurement. The outcome of a precisely controlled scientific study of an educational program is almost certain to be either inconclusive or misleading.

Logically, therefore, from Thompson's (1967) perspective, evaluation of educational programs should be conducted in the *instrumental* mode. In effect, this mode corresponds to *achievement*

monitoring. Rather than struggle with the methodological hazards associated with attempts to identify "what works," in principle, as Ebel (1980, pp. 281–292) proposes, the focus shifts to *whether* a program works. In practice, this monitoring approach already is the one most commonly adopted by school district evaluation units. For example, among the 263 districts responding to the CSE survey, 75 percent ranked as their most important task the gathering of data on pupil achievement. Task importance in this instance was measured by the proportion of the evaluation unit's time given to selected data-collection activities, as shown in Table 1.

Table 1. Percent of School District Evaluation Units Ranking Selected Activities in Order of Time Taken in 1977–78 (N=263)

Activities	Rank Order of Time Taken		
Gathering data on:	**1**	**2**	**3**
Student achievement	75%	16%	4%
Relationship between school/classroom characteristics and student achievement	2%	15%	8%
Language dominance of students	2%	12%	11%
Relationship between student socio-economic status and achievement	1%	8%	8%
Student socio-economic characteristics	0%	2%	8%
Handicapped students	1%	11%	6%
Relationship between students' race/ethnic background and achievement program	0%	3%	13%
Other	19%	33%	42%

That these data are used for *monitoring* achievement outcomes rather than for *analytic evaluation* of the instructional program is indicated by the very low proportion of districts assigning either first, second, or third priority to securing data on the relationship between school/classroom characteristics and student achievement. Further, any systematic evaluation of the effects of instruction upon student achievement must take account of student socio-economic status, the major predictor of academic performance (see Coleman, *et al.*, 1965). Despite the necessity to control for socio-

economic status in any meaningful analysis, few districts allocated it either a first, second, or third level of importance.

Confirming the findings of the CSE survey are results from David's (1978) study of local uses of Title I evaluations in 30 "exemplary" school districts across six states. David found few district evaluation units viewing the collection of standardized achievement test data as relevant to the evaluation of instructional programs in the schools. Rather, such data were seen as serving informational, or *monitoring,* needs of the state and federal levels of educational governance. Locally, the rationale for continued use of standardized tests was that they were the simplest means of responding to state and federal reporting requirements (David, 1978, p. v).

EVALUATING INSTRUCTION

Given that current activities of school district evaluation units generally cover the need to *monitor* program outcomes, providing annual data on pupil achievement relative to national norms, the problem remains of actually *evaluating* the adequacy of the instructional procedures underlying pupil learning gains. Regarding this problem of instructional evaluation David (1978) asked respondents in the 30 districts she studied how they *would* demonstrate the success of their programs. Preferred approaches, for both staff and parents, emphasized utilization of criterion-referenced, or "curriculum-embedded and other skill tests" (David, 1978, p. vi). Such tests incorporate items projected forward from the actual curriculum content. Test results thus bear directly upon the extent to which local instructional objectives are being attained, allowing teachers, the ultimate instructional decision-makers, to determine which curriculum components need more, or less, emphasis.

Devolution of responsibility for testing to the teachers at the instructional level of the school certainly respects the logic of schools as institutional organizations, with teachers conducting instructional activities relatively autonomously of the administration. To make this alternative succeed, Ebel argues,

> local school personnel need to receive training in evaluation procedures to the point where they can do the job as adequately as an external evaluator can be expected to do it. Practical evaluation techniques are not so highly technical or so remote from typical school operations as to require a separate profession of evaluation.

> People competent to design programs and to teach should also be competent to assess the results of teaching (Ebel, 1980, p. 284).

Ebel's approach addresses the problem of helping teachers to improve instructional programs on the basis of their own evaluations.

If one wishes to go beyond achievement monitoring, and also pursue program evaluation from a perspective broader than that of the individual classroom teacher, it is necessary to adopt a methodology appropriate to the loosely coupled structure of schools, drawing upon the hermeneutic rather than the experimental tradition of research design. Essentially, this involves a methodology emphasizing qualitative rather than quantitative studies of instructional programs.

QUALITATIVE EVALUATION

The alternative to what Patton (1978, p. 204) calls "the dominant hypothetic-deductive paradigm" is being applied to evaluation by persons such as Parlett and Hamilton (1976), Robert Stake (1975), and Kenneth Strike (1972). While "the natural science paradigm aims at prediction of social phenomena, the holistic-inductive, anthropological paradigm aims at understanding of social phenomena" (Patton, 1978, p. 204). This paradigm draws upon the more qualitative methods and perspectives of phenomenology, symbolic interactionism, ethnomethodology, and anthropology rather than the quantitative.

Specific objectives associated with qualitative evaluation are outlined by Parlett and Hamilton (1976, p. 144) as cited by Patton (1978, p. 209). These objectives, in relation to a school program, are likely to include:

> . . . how it operates; how it is influenced by the various school situations in which it is applied; what those directly concerned regard as its advantages and disadvantages; and how students' intellectual tasks and academic experiences are most affected. It aims to discover and document what it is like to be participating in the scheme, whether as teacher or pupil, and, in addition, to discern and discuss the innovations' most significant features, recurring concomitants, and critical processes. In short, it seeks to address and to illuminate a complex array of questions.

In practice, implementation of such "illuminative" evaluation requires specialized personnel, thus restricting its use to evaluations sponsored by state or federal agencies, rather than local districts, though the latter may well opt to encourage universities in their areas to explore this qualitative approach.

CONCLUSION

At the outset of the chapter two questions were raised: (1) What inhibits analytic evaluations of instructional programs at the school district level? (2) What might be done to facilitate more emphasis upon analytic program evaluation?

The major inhibiting factors appear tc be:

(1) conflict between the promise and the performance of national evaluations, typically based upon the experimental research paradigm;
(2) attempts to use the experimental research paradigm for evaluating instructional programs in schools, despite the fact that schools are not technically rational but institutional organizations;
(3) utilization of standardized norm-referenced tests for reporting the outcomes of externally funded programs, despite the low credibility of these tests as a means of local program assessment.

School district evaluation units tend to remove themselves from the cross-pressures inherent in the above contradictions by opting out of any sustained commitment to *program evaluation,* focusing instead upon *monitoring* program outcomes in the form of student achievement, administering tests to participants in funded programs, and to pupils at selected grade levels across all schools. Test results are reported to state education departments, to other funding agencies, and to the public locally.

While monitoring has utility in locating schooling outcomes of one district relative to another, and in identifying temporal trends, it says nothing about existing instructional procedures. To evaluate relationships between instruction and learning outcomes requires the adoption of evaluation methodologies not in contradiction with the nature of schools as institutional organizations. Such methodologies are available, as discussed here, for use by both the indi-

vidual teacher, and by external evaluators, but require commitment to criterion-referenced tests for teacher-based evaluation, and to qualitative methodologies for external evaluation.

The danger to avoid is that identified by House (1978), and discussed in more general terms by Habermas (1971). This danger is that rather than adopt evaluation approaches appropriate to the nature of schools and school processes, pressure will be exerted in the direction of making schools into technically rational organizations, in the belief that previous null findings reflect the inadequacies of schools, rather than the inappropriateness of current methods of evaluation.

References

Apsler, R. In defense of the experimental paradigm as a tool for evaluation research. *Evaluation*, 1977, *4*, 14–18.

Averch, H. A., et al. *How effective is schooling? A critical review and synthesis of research findings*. Santa Monica, CA: Rand Corp., 1972.

Bailey, S. K., & Mosher, E. K. *ESEA: The Office of Education administers a law*. Syracuse: Syracuse University Press, 1968.

Bernstein, B. *Class, codes and control*, Volume 3. *Towards a theory of educational transmissions*. Boston, Mass.: Routledge & Kegan Paul, 1975.

Bidwell, C. The school as a formal organization. In J. F. March (Ed.), *The handbook of organization*. Chicago: Rand McNally, 1965.

Campbell, D. T. Reforms as experiments. In E. L. Streuning & M. Guttentag (Eds.), *Handbook of evaluation research* (Vol. 1). Beverly Hills: Sage Publications, 1975.

Coleman, J., et al. Equality of educational opportunity. U.S. Department of Health, Education, and Welfare. Washington, D.C., U.S. Government Printing Office, 1966.

David, J. L. *Local uses of Title I evaluations*. Research Report EPRC 21. Educational Policy Research Center. Stanford Research Institute, 1978.

Dreeben, R. *The nature of teaching*. Glenview, Ill.: Scott, Foresman, 1970.

Ebel, R. L. *Practical problems in educational measurement*. Lexington, Mass.: D.C. Heath and Company, 1980.

Evans, J. W., & Anderson, J. K. Comments on an RFP for evaluation of Title I programs: Will OE learn what it wants to know? *Evaluation*, 1970, *4*, 157–160.

Goodlad, J. I. *The dynamics of educational change*. New York: McGraw-Hill, 1975.

Gorham, W. Notes of a practitioner. *The Public Interest*, 1967, Summer, 4–8.

Habermas, J. *Toward a rational society*. Boston: Beacon Press, 1970.

Habermas, J. *Knowledge and human interest*. Boston: Beacon Press, 1971.

House, E. R. Evaluation as scientific management in U.S. school reform. *Comparative Education Review*, 1978, *22*(3), 388–401.

Houston, Jr., T. R. The behavioral sciences impact-effectiveness model. In P. H. Rossi & W. Williams (Eds.), *Evaluating social programs: Theory, practice, and politics.* New York: Seminar Press, 1972.

Kerins, C. T. Accountability and the state. In E. R. House (Ed.), *School evaluation: The politics and process.* Berkeley, CA: McCutchan, 1973. Pp. 74–79.

Lortie, D. C. *Schoolteacher: A sociological study.* Chicago: University of Chicago Press, 1975.

Lyon, C. D., Doscher, L., McGranahan, P., & Williams, R. *Evaluation and school districts.* Center for the Study of Evaluation, University of California, Los Angeles, 1978.

McLaughlin, M. W. *Evaluation and reform: The elementary and secondary education act of 1965/Title I.* Cambridge, Mass.: Ballinger Publishing Company, 1975.

Metz, M. H. *Classrooms and corridors.* Berkeley, CA: University of California Press, 1978.

Meyer, J. W. *Research on school and district organization.* Paper presented at annual Sociology of Education Association Conference, San Diego, 1977.

Parlett, M., & Hamilton, D. Evaluation as illumination: A new approach to the study of innovatory programs. In G. V. Glass (Ed.), *Evaluation studies: Review annual, Vol. I.* Beverly Hills, CA: Sage Publications, 1976.

Patton, M. Q. *Utilization-focused evaluation.* Beverly Hills, CA: Sage Publications, 1978.

Riecken, H. W., & Boruch, R. F. (Eds.). *Social experimentation: A method for planning and evaluating social intervention.* New York: Academic Press, 1974.

Sarason, S. B. *The culture of the school and the problem of change.* Boston: Allyn & Bacon, 1971.

Stake, R. W. *Evaluating the arts in education: A responsive approach.* Columbus, Ohio: Charles E. Merrill, 1975.

Strike, K. Explaining and understanding: The impact of science on our concept of man. In L. J. Thomas (Ed.), *Philosophical redirection of educational research: The seventy-first yearbook of the National Society for the Study of Education.* Chicago: University of Chicago Press, 1972.

Suchman, E. A. *Evaluative research: Principles and practice in public service and social action programs.* New York: Russell Sage Foundation, 1967.

Thompson, J. D. *Organizations in action.* New York: McGraw Hill, 1967.

USCAN. Legislative history of P.L. 92-380: Education amendments of 1974. *U.S. Congressional and Administrative News,* 1974.

Weick, K. E. Educational organizations as loosely coupled systems. *Administrative Science Quarterly,* 1976, *21,* 1–19.

Wise, A. E. Why educational policies often fail: The hyperrationalization hypothesis. *Curriculum Studies,* 1977, 9(1), 43–57.

Wolcott, H. *Teachers versus technocrats.* Eugene, Oregon: University of Oregon, CASEA, 1977.

Themes and Directions

The chapters in this monograph have incorporated four author's interpretations of, or reflections on, the CSE survey data of school district evaluation units, specifically as the data relate to evaluation's role in school district operations and decision making. Because the authors are sociologists they examine school districts and their evaluation units as social entities; as one would expect, common themes run through their papers. The authors differ mainly in the emphasis they give to these common themes.

In this concluding section, we indicate three of the common themes which form the core of an argument that should be considered by school district evaluators and by evaluation researchers, especially those who are interested in increasing the power of evaluation in school districts. We will also discuss these themes in relation to some of the current work at the Center for the Study of Evaluation, and speculate on their implications for further research and development in the field of evaluation.

Theme 1: School districts are institutional organizations.

School districts are, with regard to their instructional activities, essentially institutional rather than technical organizations. An important reason for this is that schooling has a weak technical core—little is known with certainty about how teacher behavior and educational programs affect pupil learning. Thus, it is difficult to link the technical core—a fundamental feature of technical organizations—to the organizational structure. As a consequence of the institutional nature of schools, it is difficult for evaluation findings to be seen as a strong tool for school or educational program improvement. It is difficult to make credible the precise analysis of phenomena that are themselves not seen as precise.

Also, there are few incentives for school districts to become more technically oriented thereby enabling them to use evaluation data for instructional decision making. School districts have been largely successful as institutional organizations; the public has been

willing to accept symbolic representations, e.g., tenured teachers with advanced degrees, as indices that the schools have quality. Much of the technical work performed by school personnel—that is, work related to instruction and teaching—is kept shielded from public view.

Theme 2: Evaluation units have a weak position within school districts.

Evaluation units often have little power within school districts. Most such units originated in response to accountability demands from external sources, and their work in testing and evaluation tends to emphasize external "signaling." The information desired by external funding agencies, because it is designed to ensure compliance or to make comparisons, is different from that which could serve internal instructional needs. Since most evaluation units have only the time or staff to provide one kind of information, they primarily serve governmental or community clients on behalf of their districts; and they do only that monitoring and testing which appear to be required. Many evaluation units thus face a dilemma: because they do not produce much information that is valued internally, they occupy a peripheral status *vis-a-vis* other district operations. Because of this peripheral status, they are not in a position to influence district policies or operations. Nor are they given much opportunity to demonstrate what they might do in relation to the improvement of management or instruction.

Theme 3: It is difficult to produce evaluative data of utility to within-district decision makers.

The low organizational status that evaluation units have may be exacerbated by the professional stance of some evaluators. Evaluators may be seen as—and may see themselves as—having an "objective, analytical and rational" orientation. This orientation can bring them into conflict with district and school administrators who live in a world of clashing political and economic forces, and with teachers who must work in the constraints of the classroom. District administrators must often walk a tightrope between conflicting constituencies; in given situations, they may not or will not use the findings from data-based analyses of programs they are considering or operating. Put another way, school administrators may view the data from evaluation or from testing programs as only

one factor to consider in making decisions. They operate with an institutional/procedural/situation-specific rationality as distinct from a technical/method-based/generalizable rationality. Teachers may resent time taken from instruction by an evaluation they see as having little relevance to their work.

Thus, the political as well as the organizational context of a school district places severe limits on the internal utility of the work of evaluation units. The credibility of evaluators is further diminished because of the indeterminacy of cause and effect in education, as previously mentioned. Because the "technical core"—that is, certain knowledge about the causal nature of the relationship between teaching and learning—is weak, instructional recommendations based on evaluation or test findings are at best tentative. Conclusions from evaluative data can be discredited with relative ease by those who, for political, financial, economic, or professional reasons, are in disagreement with the findings.

Implications for the Future

These three themes would seem to explain why evaluative data are not now used internally, within the central office, schools or classrooms. Evaluation units, as organizational entities, face challenges with regard to integrating their "signaling" function with the production of data-based information related to the technical work of schooling.

Moreover, new policies and shifting conditions may prove to be an additional challenge for district evaluation units that seek a more stable and valued place in school district decision making. The current move away from external categorical aid programs towards block grant funding gives local educational agencies greater discretion over how and where to target external funds. The freeing up of these funds may also relax some of the specific evaluation requirements that have typically accompanied categorical grants. Given such freedom of choice, will school district administrators choose to invest their funds in supporting the work of evaluation units and redirecting the evaluation towards internal problems? In the field work CSE has recently conducted in school districts with evaluation units we have not yet found strong, broad-based support for evaluation, although there are a number of districts in which specific administrators and school board members ask for and use the evaluation units' work. Such advocates might look favorably on funding evaluations from local or internal funds.

We doubt, however, that these decisions can prevail for long when they are placed in competition against various larger political and special interest groups, e.g., unions, parents of handicapped children, who want a share of an increasingly shrinking school district budget. It may also be difficult for school district evaluation units to restructure their own thinking so as to produce the kind of reports that will capture the attention of decision makers and strengthen the resolve of administrators and Board members to retain their services. While the academic community continues to produce better data collection instruments and evaluation designs and to upgrade the quality of evaluators' training, these efforts by themselves are not likely to solve the problem of evaluation utilization within school district offices.

Thus, declining resources and considerable external political turmoil contribute to the school administrator's problems in trying to satisfy simultaneous demands for reform and for retrenchment. Powerful individual and group interests are at stake and educational administrators and policy makers may be forced towards political accommodations rather than empirical data when reaching decisions. Administrators and policy makers may not yet have come to appreciate the contribution evaluation can make to their work; they may feel that evaluations are risky, that they have potential for upsetting organizational equilibrium and so make the decision makers' tasks more difficult.

In spite of these challenges, we believe that the present situation in school districts offers new opportunities for enhancing the relevance of evaluation and testing for local decision-making purposes if properly seized by those in district leadership positions.

Communities are indeed plagued by declining sources of revenue. They therefore are pressing school administrators to demonstrate that public schools can teach children to acquire basic skills. Parents, concerned about academic and vocational opportunities for their children, often interpret their own child's performance on standardized tests as indicators of the schools' ability to teach. Simultaneously, district administrators themselves are becoming even more aware of their need to monitor both the long-term patterns of students' test scores as well as to analyze any given year's test or evaluation data in a way that suggests policy, administrative, or instructional remedies.

At the same time that school districts experience this increased press for description and diagnosis of student learning, many administrators do find that they do possess, within the district, the

capacity for responding. The past fifteen years of federal mandates for evaluation have helped equip some, if not all, districts with competent personnel who have data collection and data analysis skills which can be re-directed for internal use. If federal requirements for evaluation slacken, evaluation units may have the opportunity to justify their existence by responding to internal management needs.

Our work has led us to conclude that evaluation units have great potential for providing timely, technically sound, and useful services to school districts. They, and those in district leadership positions who believe in the value of systematically collected data, might consider some promising orientations derived from CSE's current fieldwork.

1. *Multiple definitions of evaluation units' roles.* Evaluation units can play many roles in school district functioning even though many in the past seemed limited to administering tests and monitoring compliance. Evaluation units and decision makers should explore common interests and ways in which evaluation, through new and different techniques, can serve the multiple needs of a variety of constituencies. For example, a common observation in school districts has been that evaluation units only serve those programs with external evaluation funds. Many teachers in those programs feel that such evaluations are inconsequential or even detrimental to their work. Evaluation units might well begin exploring ways in which they can serve immediate teacher needs thereby solving rather than exacerbating teacher problems. In this process, the support of central office leadership is essential. "Idea champions" who have the vision to see how evaluation findings may serve to improve teaching and learning at the classroom level can work out situation-unique solutions to the management problems of linking evaluation to internal decision making.
2. *Improved communications.* Part of evaluation's previous lack of a district constituency may have resulted from the failure of evaluators and decision makers to understand one another's world and to recognize the strengths and weaknesses of evaluation as a decision-making tool. Supportive school administrators and evaluators must now take new steps to assess the role evaluation has played and can play in district activities. Such an examination should include efforts to under-

stand the contributions and complexities of linking data with instructional change. Evaluators must become sensitive to the need to follow up their data collection activities with data dissemination activities. They must also participate in the organization problem-solving process which should follow evaluation. It seems very unlikely that school districts will spend diminishing funds on any services that are viewed as peripheral by those who control the funding decisions. Evaluators should become central and influential in the internal management of the district and in the improvement of instruction.

3. *Evaluation and instruction.* While the monograph authors have argued that education still has a weak technical base, the work of several researchers and developers (Fisher *et al.*, 1978; Stallings, 1980; Bloom, 1980; Hunter, 1971;) can, in our opinion, be seen as the foundation of an emergent technical core. If this is the case, and some school districts are beginning to identify it as such, then evaluation activities can indeed begin to contribute to understanding the relationships between processes and outcomes within given settings. School district evaluation units can work with the district's instructional program and contribute to its improvement.

We believe that the evaluation community outside of and within school districts is likely to become, in the next ten years, more intimately involved than in the past with the definition, clarification, and resolution of important educational problems. The isolation of evaluators is ending. Individuals interested in evaluation and testing are being forced to grapple with a broad range of serious concerns in education today. They are being asked to make a contribution to the knowledge base underlying learning, instruction, curriculum, and management. Evaluation and testing personnel, as they function in school districts, are likely to become more pluralistic and interdisciplinary. These individuals will come to possess a broad range of knowledge and talents. They will come to understand with sensitivity and empathy the context within which other educators work. They will come to anticipate the impact of their activities on the people who are evaluated and on those who use evaluation findings. We believe that they have the potential for making a difference.

References

Bloom, B. *All our children learning.* New York: McGraw Hill, 1980.

Fisher, C. W., Filby, N. N., Manilave, R. S., Cahan, L. S., Dishaw, M. M., Moore, J. E., and Beliner, D. *Teaching behaviors, academic learning time and student achievement: Final Report of Phase III-B, Beginning teacher evaluation study.* San Francisco: Far West Regional Laboratory, 1978.

Hunter, M. The teaching process and the learning process. *Handbook for teachers,* Siefman, E., and Alken, D. (eds.). Glenview: Scot Foresman, 1971.

Stallings, J. Allocated academic learning time revisited, or beyond time on task. *Educational Researcher,* December, 1980, pp. 11–16.