ED 205 614   TM 810 611

AUTHOR   Raizen, Senta A., Ed.; Rossi, Peter H., Ed.
TITLE   Program Evaluation in Education: When? How? To What
                Ends?
INSTITUTION   National Academy of Sciences - National Research
                Council, Washington, D.C. Assembly of Behavioral and
                Social Sciences.
SPONS AGENCY   National Inst. of Education (ED), Washington, D.C.
REPORT NO   ISBN-0-309-03143-5
PUB DATE   81
NOTE   288p.
AVAILABLE FROM National Academy Press, 2101 Constitution Avenue,
                N.W. Washington, DC 20418.

EDRS PRICE   MF01/PC12 Plus Postage.
DESCRIPTORS   *Administrative Organization: Elementary Secondary
                Education: *Evaluation Methods: Federal Programs:
                Program Effectiveness: *Program Evaluation: *Program
                Improvement: Resource Allocation
IDENTIFIERS   *Evaluation Research: *Evaluation Utilization

ABSTRACT
                In response to a provision of the Education
Amendments of 1978 concerning evaluation practices and procedures,
this report examines four aspects of evaluation in education which
focus on how funds allocated to evaluation can be spent more
effectively and yield more useful results. On this basis, groups of
recommendations are made to both Congress and the Department of
Education. These are presented in an opening summary and the
discussed more fully in each chapter. The first chapter is designed
as an introduction to the background and scope of the report. A
definition of evaluation is given in chapter two, which also
addresses congressional concern with uniform methods and measures in
the context of delineating different types of evaluation and their
appropriate use. Improvement of the quality of evaluation forms the
substance of chapter three. Discussion in chapter four centers upon
the utilization and dissemination of evaluation results. The final
chapter makes recommendations for improved management and
organization and presents implications for this derived from
preceding chapters. Appended are a glossary, references, and three
appendixes concerning types of federal educational evaluation
activities, the parties who conduct them, and the organization of the
evaluation system at state and local levels. (Author/AEF)

Program Evaluation in Education:

When? How? To What Ends?

N
R NAS
C NAE
IOM

Program Evaluation in Education:
When? How? To What Ends?

Senta A. Raizen and Peter H. Rossi,
Editors

Committee on Program Evaluation in Education
Assembly of Behavioral and Social Sciences
National Research Council

3

Committee on
Program Evaluation in Education

PETER H. ROSSI (Chair), Director, Social and Demographic
    Research Institute, University of Massachusetts
MARVIN C. ALKIN, Professor of Education and Program
    Director, Center for the Study of Evaluation,
    University of California, Los Angeles
ROBERT F. BORUCH, Director, Program on Methodology and
    Evaluation Research, Northwestern University
NATHAN S. CAPLAN, Program Director, Institute for Social
    Research, and Professor of Psychology, University of
    Michigan
CHRISTOPHER T. CROSS, Federal Marketing Manager,
    Westinghouse DataScore Systems, Westinghouse Learning
    Corporation, Washington, D.C.
EDWARD M. GRAMLICH, Director, Institute of Public Policy
    Studies, University of Michigan
FREDA M. HOLLEY, Director, Office of Research and
    Evaluation, Austin Independent School District,
    Austin, Texas
HENRY T. INGLE, Director, Program Evaluation Division,
    Office of Fair Housing and Equal Opportunity, U.S.
    Department of Housing and Urban Development
JAMES M. JONES, Director, Minority Fellowship Program,
    American Psychological Association, Washington, D.C.
SARA LAWRENCE LIGHTFOOT, Professor of Education and
    Social Policy, Graduate School of Education, Harvard
    University
WILLIAM C. ROCK, Professor, Department of Educational
    Administration, State University of New York, Brockport
HOWARD ROSEN, Adjunct Professor, School of Business
    Administration, American University

iii

KARL E. WEICK, Nicholas H. Noyes Professor of
   Organizational Behavior and Professor of Psychology,
   Graduate School of Business and Public Administration,
   Cornell University
ROBERT K. YIN, Visiting Associate Professor, Department
   of Urban Studies and Planning, Massachusetts Institute
   of Technology

SENTA A. RAIZEN, Study Director
DIANE L. GOLDMAN, Administrative Secretary

Contents

Preface

The Education Amendments of 1978 (P.L. 95-561), which
reauthorized the major federal elementary and secondary
school programs, included the following provision:

STUDY OF EVALUATION PRACTICES AND PROCEDURES

SEC. 1526. The Commissioner of Education
shall conduct a study of evaluation practices
and procedures at the national, State, and
local levels with respect to federally funded
elementary and secondary educational programs
and shall include in the first annual report to
Congress submitted more than one year after the
date of enactment of this Act proposals and
recommendations for the revision of
modification of any part or all of such
practices and procedures. Such proposals and
recommendations shall include provisions

(1) to ensure that evaluations are based
on uniform methods and measurements;
(2) to ensure the integrity and
independence of the evaluation process;
and
(3) to ensure appropriate follow-up on
the evaluations that are conducted.

This requirement has provided the impetus for the
present report. In response to the legislative request,
the National Academy of Sciences was asked by the Office

of Education (OE) to undertake a study of program
evaluation in education. The purpose of the study was to
recommend ways of increasing the effectiveness and
usefulness of the OE's evaluation efforts. The study was
started late in 1979 and completed under the auspices of
the new Department of Education, the successor agency to
OE.

It was explicit in the request made by OE that the
core of the study would be a report by an expert
committee. The Committee on Program Evaluation in
Education came to life in early 1980, convened under the
auspices of the Assembly of Behavioral and Social
Sciences. Its membership was selected to represent
appropriate disciplines as well as different viewpoints
and responsibilities regarding evaluation, in recognition
of the fact that the problems to be addressed related as
much to the organization, management, and policy uses of
evaluation as to questions of evaluation strategy,
methodology, and quality. The disciplines represented on
the Committee included communications, economics,
educational administration, educational psychology,
experimental psychology, political science, social
psychology, sociology, sociology of education, and
statistics (psychometrics). The experience represented
included: carrying out large-scale and smaller
evaluations in different settings (university, local
school system, private sector); commissioning evaluations
and managing more general programs of support for applied
social research and development (R&D) within several
government agencies; serving as staff to a major
congressional education committee; and carrying out
pertinent research on methodology and utilization of
evaluations and on social R&D. Several members had also
conducted general assessments of the field of evaluation.

The Committee held three two-day meetings and a longer
working conference to develop the substance of the
report. Richard A. Berk of the University of California,
Santa Barbara, assisted the Committee as a consultant
during the working conference. During its first two
meetings, the Committee focused on defining the key
issues to be addressed. Senior staff from the Department
of Education and from education committees in Congress
met with the Committee to give us the benefit of their
views. (See Appendix D for a list of participants.) In
addition to the concerns expressed by Congress with
methods, integrity, and follow-up, Department officials
asked that the following organizational topics be

viii

addressed: location of evaluation activities within the
Department, coordination of evaluation within the
Department, participation in evaluation design and use by
program and planning officials, and continuing advisory
mechanisms for evaluation. Department staff also raised
the following nonorganizational issues: distinguishing
among types of evaluations, planning of evaluations,
strategic considerations in evaluation management, and
appropriate utilization.

Starting from those expressed concerns, the Committee
explored other related issues and came to organize the
report around four major topic areas: distinguishing
between evaluation types and choosing appropriate
strategies and procedures; improving the quality of
evaluations; increasing the effective use of evaluations;
and improving the organization and management of
federally funded evaluations in education. The
congressional concern with uniform methods and measures
was subsumed under the broader topic of evaluation
strategies and procedures, since consideration of methods
and measures is possible only in the context of a
specific set of policy questions and after an evaluation
strategy and procedure have been determined.

In carrying out its study, the Committee relied on
various kinds of information to supplement the members'
knowledge and experience. Members and staff conducted
informal interviews with employees and ex-employees of
OE, of the Department of Education, of other federal R&D
support agencies, and with congressional staff familiar
with the provision calling for the assessment of
evaluation practices. (For a list of persons
interviewed, see Appendix D.) Two papers were
commissioned from consultants to supply detailed
information on the evaluation activities within the
Department and on the performer communities that carry
out evaluation studies; they appear as Appendixes A and
B. A third paper, contributed by Committee member Freda
M. Holley, provided insight into evaluation activities at
the state and local levels and is included as Appendix
C. Working papers were also prepared by me and
Richard A. Berk and by members Marvin C. Alkin, Robert F.
Boruch, and Robert K. Yin. These have been published by
their individual authors under the aegis of the Center
for the Study of Evaluation (Baker 1980). Material from
these papers and from various drafts of chapter sections
prepared by other Committee members has been incorporated
in the report. Additional background material available

to the Committee included agency planning documents,
annual reports, and internal critiques relating to
evaluation activities and their application to decisions
about programs.

This report is not a comprehensive examination of
program evaluation in education. The intent of the
sponsoring agency was to have a group of experts apply
their knowledge and experience to the problems identified
by Congress and the Department. This has structured both
the selection of subject matter and the nature of the
evidentiary base, which is drawn largely from existing
data and analyses. Neither money nor time was available
for an empirical study, such as: an examination of the
quality of procurement instruments, of resulting
proposals, or of evaluation reports; systematic surveys
of sponsors or performers on their experience with
different types of evaluations and management practices;
or primary analysis of the use of evaluation results.
However, the Committee was able to use the findings of a
second and more extensive project funded by OE in
response to a congressional request. This project,
located at Northwestern University, included collection
of empirical data and examination of the literature on
evaluation of federally supported education programs at
the national, state, and local levels. During its third
meeting, the Committee reviewed the reports of this
project and became familiar with its findings (Boruch and
Cordray 1980). In addition, the director of the
Northwestern project served on the Committee, which was
thus able to take advantage of the complementary nature
of the two projects.

The Committee is grateful for the assistance it
received from many other sources. We owe special thanks
to John W. Evans, the former head of the central
evaluation unit of the Department of Education, who made
himself and his staff fully available to the Committee,
and to Marshall Smith, former executive assistant to the
Secretary. They and other staff within the Department of
Education provided much data and were generous with their
time and the effort needed to comply with our requests
for material and information. Staff members from the
National Science Foundation and from Congress also gave
generously of their time.

Members of the Assembly of Behavioral and Social
Sciences (ABASS) of the National Research Council and of
the Report Review Committee of the National Academy of
Sciences provided thoughtful comments on an earlier draft

of the report that helped improve the final version. We are grateful, too, to David A. Goslin, executive director of ABASS, for his support and valuable suggestions, to Eugenia Grohman, associate director for reports of ABASS, who critically edited the report, and to Elaine McGarraugh, editorial assistant, who supervised its production.

Finally, we wish to thank Rose S. Kaufman, whose administrative support early on facilitated the organization and first meetings of our Committee, and Diane L. Goldman, who ably took over from her as our administrative secretary, typed the many versions of the report, and provided us with much needed logistical support and technical assistance.

Peter H. Rossi, Chair
Committee on Program Evaluation in Education

xi

Summary

Evaluation as an established field of applied social
science research has grown rapidly over the last 20
years, accompanied by the expectation that the empirical
knowledge resulting from evaluation studies would improve
the process of making decisions about social programs.
In education, more than $40 million is now spent per year
for evaluation activities by the Department of Education;
about $60 million more in federal funds is spent by other
federal agencies and by state and local agencies. But as
the number of evaluation studies and their sophistication
have grown, so has concern that evaluation work has not
lived up to its potential. In response to such concerns
on the part of Congress, the Committee on Program
Evaluation in Education examined four aspects of
evaluation in education: the varieties of evaluation and
their respective roles; the quality of evaluation
efforts; the use of evaluation results; and the
organization and management of evaluation activities. We
focused on these topics because they were identified to
be of greatest interest to the two primary audiences for
our report: members of Congress and their staffs and
high-level officials in the Department of Education.

FINDINGS AND CONCLUSIONS

Two major findings permeate the Committee's report.
First, evaluation must be viewed as a system that
involves many organizations and many parties. Attempts
to improve the quality of evaluation studies or to
increase the use of evaluation results must deal with

1

13

systemic problems rather than with the specific
shortcomings of any individual evaluation. Therefore,
much of this report deals with such systemic issues as
the role of evaluation, the context in which it takes
place, and the diverse interests of the many groups
concerned with federal education programs. Second, both
the quality and the use of evaluations could be
considerably enhanced through better management
procedures. At present, the processes for soliciting and
funding studies constrain creativity; quality controls
are insufficient; limited review procedures at all stages
inhibit the development of an active intellectual
marketplace--the most effective arbiter of quality and
use. Hence, most of our recommendations are designed to
improve the procedures that now govern federally funded
evaluations in education. Improvement in management
procedures is the single most important step that
Congress and the Department could take if they wish to
achieve better quality in evaluations and to increase the
likelihood that evaluation results will be used
appropriately.

The Role of Evaluation

To understand what evaluation can contribute to the
making of policy, one must understand its limited role in
affecting decisions that are largely shaped by other
forces. In any political decision, many parties with
diverse interests are likely to have a stake, and
evaluators are often asked to respond to several
audiences and competing constituencies. Even though
evaluations are frequently conducted at the behest of
governmental authorities making decisions about programs,
other audiences will respond to evaluation information as
well and use or not use it as it furthers their
objectives. Different audiences have need for different
types of information; different policy issues require
different types of studies. Unless the policy questions
to be addressed are clear to those who ask for
evaluations and to those who carry them out, the
perception that much evaluation work is irrelevant to the
policy process is likely to persist.
   The diversity of research activities all going under
the general name of evaluation has led to considerable
misunderstanding. The diversity has come about because
it has become evident that studying the effectiveness of

operating programs--the traditional focus of
evaluation--does not answer some important questions;
research is also needed in planning and implementing
programs. During the planning phase, there are questions
of need and how to meet those needs. Survey and
ethnographic studies can establish the extent and
distribution of an educational problem; controlled pilot
testing and field tests can determine the effectiveness
and feasibility of alternative interventions for
relieving the problem; and economic analyses can be used
to make cost estimates. Once a program is established
and operating, there are questions of fiscal and coverage
accountability. Analyses of administrative records can
determine whether funds are being used properly and
whether the program is reaching the intended
beneficiaries, although supplementary fiscal audits and
beneficiary studies are sometimes required. Finding out
whether the program is being implemented appropriately
requires, in addition to program administrative records,
special surveys of program services and ethnographic
studies. Finally, there are questions of program impact;
they can be addressed definitively only through rigorous
and often costly research methods. Consequently impact
evaluation should be undertaken only if the requisite
skills and resources are available.

    Not all programs can be fully evaluated: that is, not
all questions can be answered for all programs. In
particular, meaningful impact evaluation is possible only
for programs for which intended beneficiaries and effects
can be clearly specified. There are two kinds of
programs for which such specification is extremely
difficult or impossible. For a program having vague
goals or many diverse goals, evaluators and those who
commission an evaluation must be able to agree on which
goal should be assessed and whether appropriate measures
are available to assess it. For a program in which local
sites are given autonomy to develop their own specific
objectives and means of reaching them, one cannot
evaluate for national impact by aggregating effects over
many diverse sites (though the effectiveness of
individual local projects may be evaluated). General
judgments about a national program become possible over
time, however, as knowledge from studies of individual
sites accumulates.

    In an effort to increase the quality of information
furnished through local evaluations, Congress has sought
to encourage uniformity of methods and measurements in

evaluation. At this time, the Committee does not
consider such uniformity an appropriate means for
controlling quality, since requiring uniformity may
prematurely inhibit further advances in methodology.
Instead, evaluation methods should be subjected to the
full test of the intellectual marketplace through
intensive review and critique.

Improving the Quality of Evaluations

The few systematic or informal surveys of evaluation
studies in education give some credence to the frequently
voiced dissatisfaction with the general level of their
quality. There appear to be several reasons that the
quality of evaluations in education has been found
wanting. First, the unrealistic expectation that
complicated evaluation issues can be addressed by a wide
variety of agencies has led to some inappropriate
assignments of evaluation responsibility. For example,
only a few large and sophisticated school systems and a
handful of states have the capacity to carry out rigorous
studies of program impact. In addition, the objectivity
that is necessary for good evaluation is sometimes
compromised at the state and local levels because much of
the evaluation funding, though supplied by the federal
government, is controlled by local program managers or
state administrators. Evaluation requirements imposed on
local and state authorities should match their
capabilities, and fiscal and organizational arrangements
should foster the integrity of local and state studies.
   A second reason for the low quality of evaluations
arises from the way in which federal evaluation
activities in education are managed. Though the amount
of money spent on evaluation represents only about 0.5
percent of the total federal support for education, it is
a major source of income for private-sector research
firms; moreover, evaluation work is heavily concentrated
among the larger of those firms. This concentration has
come about because of the current procedures for
sponsoring and carrying out evaluations. Procurement
documents are highly complex and often include detailed
specifications on the various technical aspects of
evaluation. Internal planning procedures and design of
requests for proposals (RFPs) take so long that little
time is left for response. Universities, minority firms,
and small businesses, unlike large firms, are unable or

unwilling to compete under such conditions. The lack of diversity among evaluation contractors reduces the possibility of new ideas entering the evaluation system and thereby improving it. Perspectives of beneficiary populations, in particular, are underrepresented on both the sponsor and the performer sides.

Flexibility in evaluation, which could contribute to quality, has also been reduced because of emphasis in the past on large studies. The restrictions on creativity imposed by this approach are aggravated when a single individual or small group within the Department develops the main procurement instrument, as is usually the case. An additional constraint on flexibility and creativity is the current monitoring process, which makes it difficult to adjust the course of a study because of changed field conditions or because a different research direction is warranted.

A third explanation for problems of quality is that the intellectual marketplace for appraisal and scrutiny of evaluations has yet to be fully formed. Generally, there is no review by outside experts during the procurement phase when the main elements of a study are being designed; the lack of diversity among competitors for evaluation work further inhibits opportunities for the marketplace to operate; and, upon completion of a study, external review of final reports happens only sporadically. Institutional mechanisms for encouraging ample discussion by experts and parties at interest of plans for and findings of major studies are spotty at the federal level; they are largely absent at the state and local levels.

Using the Results of Evaluation

A frequently voiced criticism of evaluation is that evaluation findings are seldom used. Implicit in this criticism is the notion that utilization means direct and often immediate changes in policy and program. In fact, there are several different types of utilization, not all immediately apparent. Moreover, the dissemination of findings does not automatically lead to utilization, nor is utilization synonymous with change.

Evaluation findings may be used for making specific changes at a given time, as commonly envisaged in discussions of utilization. Findings may also be used to confirm that changes are not needed. But information may

also be considered and not used because it is
inappropriate or because the indicated directions for
policy are infeasible. Moreover, even when there is no
immediately discernible use of knowledge derived from
evaluations, it cumulates over time and is slowly
absorbed, eventually leading to changes in concepts and
decision perspectives.

There are important limits to the use of evaluation
results in the short run. Social problem solving is and
ought to be a political process; the forces and events
impinging on decisions about programs are often more
powerful than empirically derived evidence. The
environment in which decisions are made seldom permits
swift and unilateral action; new information may actually
slow down the process, since it may make decisions more
complicated. For these reasons, while evaluators and
sponsors should do their best to disseminate evaluation
findings, they cannot ensure utilization.

Dissemination can be improved in a number of ways,
however. At the very least, evaluation results must be
communicated to the primary audience. Copies of reports
must be available; primary data should be accessible for
reanalysis. Unfortunately, none of these minimal
dissemination steps is now routine. Assuming that
information is made available, other important factors
affecting its use include whether it is perceived to be
objective and whether it is structured and reported in a
way that is relevant to potential users. Timeliness is
also important, particularly when direct application to
specific decisions is intended.

Because evaluation results are more likely to be used
when they address issues of importance to specific
audiences, concern with the use of evaluation findings
cannot begin when final reports are ready to be
disseminated. The primary audience and its information
needs of a given evaluation should be identified at the
inception of the study. Such initial identification will
help define the type of evaluation to be undertaken, the
issues to be addressed, the sort of information to be
collected, and the form of reporting and communication
that is likely to be most effective. The language of
evaluation reports is often a barrier to use: reports
must be intelligible to the intended audience(s) and
should be augmented by more informal means of
communication, including person-to-person interpretation
of results. Linking mechanisms that mediate between
researcher and audience can facilitate the spread of
knowledge and the utilization process.

Organizing and Managing Evaluation Activities

The Department of Education has accountability and
oversight responsibilities with regard to federal
education programs and must carry out evaluation
activities that address those responsibilities. The
Department should also develop knowledge about programs
that can be used to improve both their management and
their contribution to more effective education. Finally,
the Department should be able to formulate new programs
based on tested alternatives that speak to unmet needs in
education.

At present, evaluation responsibilities are assigned
to several different units within the Department, and to
state and local agencies. Fiscal audits and
investigations on compliance with civil rights laws are
appropriately carried out by offices created specifically
for these functions. Similarly, local and state agencies
are appropriately responsible for supplying fiscal and
beneficiary information needed to administer federal
programs. However, the assignment of other types of
evaluation responsibilities among levels of government
and within the Department varies remarkably from program
to program, despite the existence of a central evaluation
unit.

Though some decentralization of activities is
appropriate, assignment of responsibilities should be on
a more systematic and purposeful basis. The Committee
suggests the following guidelines:

* Collection of information on beneficiaries served
and on allocation of resources should continue to be a
requirement for state and local agencies. When agencies
do not have adequate capability for accurate reporting,
technical assistance ought to be provided. An important
caveat is that reporting requirements should not generate
more information than can be digested at the level
(federal or state) receiving the reports. No requirement
should be imposed on all state and local agencies that
goes beyond the basic reporting needed for accountability
functions, such as studies of program effects and
cost-effectiveness analyses. Such studies should be done
on a national sample basis or by selected local or state
agencies of proven competence and with sufficient
resources.

• The Inspector General should continue to have
responsibility for fiscal audits. Coverage of
beneficiaries and program delivery should be monitored by
the officials who administer programs at the federal
level, but the central evaluation unit should, from time
to time, run independent studies as checks. As its major
responsibilities, the central evaluation unit should, in
cooperation with the program units, carry out studies to
establish whether and how specific programs can be
evaluated, sponsor documentation of program process and
implementation, and support studies aimed at the
improvement of existing programs or the development of
new ones. The research office of the Department should
help administer grant programs for evaluation studies and
support research on the methods and processes of
evaluation.

Decentralizing evaluation responsibilities to any
degree creates the problem of how evaluation dollars can
be used effectively when they are dispersed among three
levels of government and among many of the Department's
units. First, adequate reporting of evaluation
activities and expenditures must be instituted at all
levels and for all units. Second, the central evaluation
unit should be responsible for the coordination of
evaluation throughout the Department, particularly with
respect to planning and reporting procedures. The unit
should also provide technical assistance and review for
the design and procurement of individual studies done by
other units, and it should be responsible for a
systematic process of review of interim and final reports
by inside and outside experts. A special dissemination
branch within the central unit should help other offices
with dissemination of findings from evaluation studies. ·
    The central evaluation unit will not be able to carry
out effectively the suggested evaluation and coordination
responsibilities as long as it is subsumed within the
management arm of the Department. The implicit message
of this organization is that only the management
perspective of evaluation is important. The Committee
believes that evaluation must address the substance of
policies and programs, not only their management.
Therefore, administrative arrangements should be changed
so as to give top decision makers within the Department
more direct access to the central evaluation unit.

RECOMMENDATIONS

The Committee has two sets of recommendations, one for
Congress and one for the Department. The recommendations
are presented and the discussion of them summarized in
the following two sections; the chapter numbers in
parentheses indicate where the more detailed discussions,
are found.

Recommendations to Congress

The first recommendation to Congress is concerned with
obtaining a better match between the information that
results from evaluation studies and the information that
is useful in making decisions about programs. The next
three recommendations, C-2, C-3, and C-4, are intended to
improve oversight and accountability for evaluations
carried out with funds from federal education programs.
The last recommendation to Congress addresses management
constraints external to the Department.

**Recommendation C-1. When Congress requests evaluations,
it should identify the kind of question(s) to be
addressed. (Chapter 2)**

   Given the diversity of evaluation activities
misunderstandings about what information is needed have
frequently arisen between Congress and the Department and
its evaluation contractors. Congress should attempt to
make more explicit whether it needs information about
program services, about program coverage, about program
impact, or about other program aspects. Such clarity
will make it more likely that useful information will be
delivered as a result of an evaluation effort. The
primary audience(s) for the results of the requested
evaluations should also be identified, since different
audiences need different types of information.
   Clarity of congressional intent can be brought about
in two ways. When specificity about questions and
audiences is not possible ahead of time, evaluation staff
within the Department need to engage in a continuing
dialogue with members of Congress and their staffs to
refine the policy issues to be addressed. Alternatively,
legislative language can specify such issues when
Congress wants specific information. Legislative

21

language regarding evaluation should refrain, however,
from specifying details of research method (such as
sampling procedure or use of control groups) or of
measurement. The choice of methods depends in part on
specific evaluation conditions and contexts and should be
done by technical experts only after careful
consideration of all facets of an evaluation.

Recommen.ation C-2. Congress should separate funding for
evaluations conducted at the state and local levels from
program and administrative funds. (Chapter 3)

   Under present circumstances, the amount of money
invested and the kind of evaluation done at the state and
local levels is, in too many instances, controlled by
those who administer and run programs. This puts the
quality and integrity of state and local evaluation
activities in jeopardy. Moreover, the current
arrangement makes it impossible to know how much of the
federal funds potentially available for evaluation are
actually used for that purpose. Congress may also wish
to consider a percentage set-aside for evaluation of
programs at the state and local levels, as is now
legislated for a number of programs at the national level.

Recommendation C-3. Congress should institute a
diversified strategy of evaluation at the state and local
levels that would impose minimum monitoring and
compliance requirements on all agencies receiving federal
funds but allow only the most competent to carry out
complex evaluation tasks. (Chapter 3)

   All state and local agencies receiving federal funds
for education programs should be required to provide an
accounting of the distribution of funds and of
beneficiary coverage for each program. When specific
services and procedures are mandated, these should also
be subject to reports to ensure compliance. The Congress
should require the Department to institute appropriate
quality control procedures to raise the quality of state
and local data. Evaluation tasks that go beyond
accountability questions, however, should only be
required of state and local units on a highly selective
basis. Congress may wish to consider authorizing a
competitive grants program, possibly administered through

the National Institute of Education, for school systems
and states that would provide for funding a few of the
most technically promising proposals for impact
assessments of local programs or for program improvement
based on evaluation of alternative program strategies.

**Recommendation C-4. Congress should require an annual
report from the Department of Education on all evaluation
expenditures and activities.** (Chapter 3)

The annual evaluation report currently required from
the Department should be expanded to cover all federally
funded evaluation activities in education, including all
of those in the Department as well as those carried out
by state and local agencies. Expenditures at all levels
should be specified; activities, findings, and their use
should be briefly described.

**Recommendation C-5. Congress should authorize a study
group to analyze the combined effects of the legislative
provisions and executive regulations that control
federally funded applied research.** (Chapter 5)

One of the causes of the lack of timeliness and
relevance of evaluation studies is the accumulation of
rules and regulations governing the whole process of
funding and carrying out applied research in the social
service area. While almost every provision now on the
books or enforced through executive practice is there to
provide some safeguard and may be reasonable when
considered in isolation, in the aggregate they have
negative effects. The trade-offs between the benefits of
the safeguards and the obstacles they create against
producing timely and relevant applied research at
reasonable cost deserve careful scrutiny. Simplification
and reform may be in order.

Recommendations to the Department of Education

The recommendations to the Department concentrate on
management issues for two reasons. First, as noted, we
believe that the quality of evaluations could be
considerably improved and the use of evaluation findings
increased through better management procedures. Second,

the Department has the power to change many of its
current operating procedures, while it may be able to do
relatively little about such external constraints as the
development of the evaluation field, the size of its
budget, or agency personnel ceilings. The
recommendations on procedures are organized into those
intended to develop better strategies for overall
evaluation planning within the Department and for
planning individual studies; those intended to increase
the quality of evaluations, including three on training
and technical assistance; and those intended to
facilitate use. The last three recommendations speak to
improvements needed in general management procedures.

On Evaluation Strategy

**Recommendation D-1. In evaluations initiated by the
Department of Education, the kinds of evaluation
activities to be carried out should be specified clearly
and should be justified in terms of program development
or program implementation.** (Chapter 2)

This recommendation is analogous to Recommendation C-1
to Congress. It emphasizes the need to think through
what type of evaluation activity is appropriate to any
given stage of planning or implementation of a proposed
program or an existing program. For example, top-level
Department officials need to specify what they wish to
know about a program, why they wish to know it at some
specified time, and what audiences other than themselves
have information needs that must be satisfied through
evaluation activities. All these needs must be
coordinated with legislated requests for evaluation.
(See also Recommendation D-10 on planning.)

**Recommendation D-2. When pilot tests of proposed major
programs are conducted, pilot tests of evaluation
requirements should be conducted simultaneously to
determine their feasibility and appropriateness.**
(Chapter 2)

While pilot tests of a program are being made, it is a
relatively easy matter to pilot-test the proposed
evaluation. Such a pilot test can be used to find out
what measurements can and cannot be made of program

13

benefits, how programs should account for and measure
costs, which testing instruments and procedures are
disruptive and which are not, how large a sample of
beneficiaries is needed to get valid program
measurements, and so forth. If a pilot test of an
evaluation were carried out in conjunction with the pilot
test of a program, the design of both the program and of
the evaluation requirements would be strengthened.

Recommendation D-3. The National Institute of Education
should continue and strengthen its program of support for
research in evaluation methods and processes. (Chapter 2)

   The advances made in the technical aspects of
evaluation have been considerable, but uneven. The
Committee believes that too much attention has been given
to investigating problems in the use of randomized
controlled experiments. Other important problems in
methodology have not received sufficient attention, for
example, methods for studying the delivery of services,
for investigating the properties of achievement tests
when used in the evaluation of programs, and for
assessing the impact of programs that cannot be studied
through the usual experimental paradigms. Another
neglected area of research is the process of evaluation
itself: how studies are commissioned and initiated, how
they are managed, what laws and procedures impinge upon
them. The Committee's work indicates that current
procedures constrain the quality and the use of
evaluations, but how these processes operate is poorly
understood; therefore, it is difficult to design
effective remedies.

On Quality, Training, and Technical Assistance

Recommendation D-4. The Department of Education should
provide funds for training programs in evaluation to
increase the skills of individuals currently charged with
carrying out or using evaluations and to increase the
participation of minorities. (Chapter 3)

   The field of evaluation has grown more rapidly than
the pool of skilled evaluators. As a consequence, there
are many people working as evaluators whose training has
been haphazard and inadvertent and who may not be fully

familiar with more recent advances in techniques and
methods. Others may lack adequate knowledge of the
educational system or of the special needs of the groups
to be helped by federal education programs.

A primary training need concerns the
underrepresentation of minority group members in the
educational evaluation enterprise. Well over half of all
education programs target minority group persons as
recipients of services. The Committee believes that the
quality of evaluation would be improved by the employment
of minority persons who are also well trained
technically. For example, intimate personal knowledge of
the circumstances of beneficiaries will help to define
outcome measures that are more relevant to beneficiaries
and more closely related to improving the effectiveness
of programs. Hence, we believe that such perspectives
should be represented to the fullest extent possible in
the evaluation of such programs. Fellowships and
internship programs in evaluation that include specific
priorities for minority group persons would be doubly
valuable; they would produce good researchers and they
would enrich the evaluation system.

A second concern related to training is the
relationship between the evaluator and the administrator
or educator. The communication gap be   een the two that
inhibits the use of evaluation may be   rrowed by
appropriate training on both sides.   tives and
program staff would benefit from gre.   nowledge of the
language of evaluation and how evaluat n   ght be used;
evaluators need exposure to the problems, procedures, and
constraints of federal education programs. Evaluators
also need to improve interpersonal and communications
skills in order to convey evaluation information
effectively.

Technical training for evaluation staff is also
necessary, both within the federal government and at the
state and local levels. There have never been sufficient
numbers of staff trained in either rigorous evaluation
methods or in research, and there have been rapid
developments in the field. Evaluation is currently
practiced by those from almost every type of background
possible, including many with no more preparation than
that of classroom teaching. Practicing evaluators need
opportunities to upgrade and improve their skills. As
one way of meeting this need, the Department should
consider funding short-term institutes and conferences
providing up-to-date knowledge to the evaluation

26

community. (See also Recommendation D-17 on training opportunities for federal staff.)

Recommendation D-5. The Department of Education should structure the procurement and funding procedures for evaluations so as to permit more creative evaluation work by opening up the process and allowing a period for exploratory research. (Chapter 3)

The more complex the evaluation, the less likely is it that one can spell out ahead of time the best methods for addressing the questions that the evaluation is designed to answer. The current RFP process in particular ignores this fact. The Committee believes that the RFP process can be made more flexible. RFPs for large studies should include a period of exploratory research; they should also provide for side studies that address questions integral to the evaluation that emerge after it is under way. Proposers should be given the freedom to specify alternative methods and to suggest side studies. Most important, sufficient time for developing proposals must be allowed.

Mechanisms other than RFPs for funding evaluations can also be used to open up the system. For example, unsolicited and solicited proposals, 8-A contracting, cooperative agreements, basic ordering agreements, and grant awards are each appropriate to given evaluation tasks. The Committee's recommendation that a greater variety of funding methods be employed does not imply that the use of RFPs be drastically reduced. Flexibility in the award process, we believe, will permit the introduction of new ideas that may contribute to higher-quality evaluations. Flexibility will also allow greater participation by minority organizations and researchers.

Recommendation D-6. All major national evaluations should be reviewed by independent groups at the design, award, and final report stages. Review groups should include representatives of minorities and other consumers as well as technical experts. The results of their review should be made broadly available. (Chapter 3)

This recommendation also is intended to open up the process. There are three facets to it: improving the

technical quality of evaluations, assuring early
contribution and involvement from those most affected by
programs (beneficiary groups, teachers, etc.), and making
use of findings more likely through public exposure and
understanding.

When the RFP process is used, the agency itself should
solicit as much outside advice as possible, through
development of concept papers, planning conferences, and
other pre-RFP activities. Proposal evaluation and
selection procedures should include experts from outside
the sponsoring agency. After award of a contract, the
contractor also should solicit the views of outsiders.
Then, when the project is done, outsiders should again
review the work, its assumptions, its technical
ambiguities, and its policy implications. Reviews of
completed work should be widely disseminated in order to
encourage discussions cf the findings. The Department
might sponsor an annual conference on important
evaluations that are at various points--design,
completion of final report, reanalysis. If this were
done, the educational community would know where to look
for the latest evaluation results and criticisms and be
apprised of impending work.

Recommendation D-7. All statistical data generated by
major evaluations should be made readily available for
independent analysis after identifying information on
individual respondents has been deleted. (Chapter 3)

When possible, ethnographic data and case study
material, similarly treated to protect privacy and
confidentiality, should also be made available.

Making primary data from evaluations available will
require support in major evaluation contracts for
documentation, storage, and dissemination of data and the
creation of explicit agency policy on access to data.
Since the objective is to generate adequate examination
of the methods and findings of major evaluation studies,
independent review and reanalysis should be supported by
the Department as part of its evaluation and research
programs.

Recommendation D-8. The Department of Education should
explore alternative approaches to technical assistance
for state and local evaluation needs. (Chapter 3)

The technical assistance needs of state and local
agencies are not uniform. They vary with the size of the
agency, the sophistication of the agency's evaluation
staff, and with the complexity of the federal program
activity in the agency. The technical assistance centers
associated with Title I are one approach to meeting such
needs. Another approach would be to identify or fund
exemplary models of monitoring and reporting and to
disseminate the procedures involved. A third approach
would be to develop the capability of state agencies to
provide technical assistance to less sophisticated local
agencies.
    Technical assistance should also cover organizational
and personnel issues. In particular, state and local
agencies need to be aware of the desirability of
separating an evaluation unit from program administration
in order to avoid conflicts of interest. Work already
done by some state and local agencies on optimal
institutional arrangements, personnel requirements, and
procurement policies for extramural work can form the
basis of advice and assistance to others. (See also
Recommendation D-16 on minimum requirements for
monitoring and compliance reporting.)

On Utilization

Recommendation D-9. The Department of Education should
test various mechanisms for providing linkage between
evaluators and potential users. (Chapter 4)

The Department should consider establishing a unit
charged with studying, developing, and instituting
knowledge transfer mechanisms and evaluating their
effectiveness. Alternatively, outside experts might be
charged with this responsibility. Appropriate activities
would include assessing proposed dissemination plans,
performing needed translations of evaluation reports,
funding research on the communication and use of
evaluation information, and developing procedures
designed to improve the day-to-day use of evaluation
data, at least within the Department.

Recommendation D-10. The Department of Education should
institute a flexible planning system for evaluations of
federal education programs. (Chapter 4)

A workable planning system must provide for
appropriate information to be available for recurring
legislative decision cycles on education programs; it
must accommodate an ongoing program of evaluation studies
addressing problems that are poorly understood, and it
must be sufficiently flexible to allow response to
interesting but unanticipated questions that arise as a
result of ongoing research, changes in policy, or
development of new programs. The evaluation plan for any
major education program should contain a series of linked
studies, some of which furnish factual information in
reasonably short time and some of which address issues of
long-term interest.
   Although planning does not necessarily lead to an
agenda that is subsequently carried out in detail,
planning almost always leads to an improved sense of
priorities, provides a forum in which competing interests
can reach accommodations, and induces an active as
opposed to a reactive stance toward essential activities.

Recommendation D-11. The Department of Education should
establish a quick-response capability to address critical
but unanticipated evaluation questions. (Chapter 4)

In order to be fully responsive to the information
needs of its primary audiences, the Department must be
able to combine a deliberative planning process that
allows time for field and constituency involvement with a
quick-response capability that can address unanticipated
but critical evaluation questions as they arise.
Department staff charged with evaluation responsibilities
should be able to respond within 2-6 months to
evaluation-related questions to which Congress or
top-level Department officials seek prompt answers.
Several extramural mechanisms are available for this
purpose, for example, maintaining lists of prequalified
contractors who can be given specific task orders on
short notice or using 8-A contracts and awards to
SBA-eligible firms.

30

Recommendation D-12. The Department of Education should
ensure that evaluations deal with topics that are
relevant to the likely users. (Chapter 4)

   In order to increase the relevance of evaluation
results, primary audience(s) must be specified prior to
the beginning of a study. When conditions change during
the course of a study that might affect the usability of
the findings, study objectives and design should be
reconsidered to ensure that the study will remain
relevant. Efforts should be made to deliver reports on
time, especially when study results are intended for
decisions that are made at specified times.

Recommendation D-13. The Department of Education should
ensure that dissemination of evaluation results achieves
adequate coverage. (Chapter 4)

   All RFPs and grant announcements should include
requirements for a dissemination plan oriented t' 'ard
utilization, and proposal evaluation should give
appropriate weight to the quality of the proposed
dissemination plan. Dissemination plans should include
specification of audiences and their information needs,
strategies for reaching the audiences, provision for an
adequate number of report copies and other materials, and
mechanisms for adapting the dissemination plan as the
study proceeds. Budget negotiations should recognize
that adequate dissemination is costly and cannot be an
afterthought.

Recommendation D-14. The Department of Education should
observe the rights of any parties at interest and the
public in general to information generated about public
programs. (Chapter 4)

   Findings from evaluations must be made available to
those who are importantly affected by the programs being
evaluated, including those who manage them, those who
provide program services, and those who are intended to
benefit (or their representatives). Since evaluations
are paid for with public funds, they should also be made
available to the public at large. The Committee is aware
of the dangers in providing too much autonomy to
evaluation units and contractors, but public interest

needs suggest that, at the dissemination stage,
evaluators should be guaranteed a certain degree of
autonomy. Appropriate changes should be made in contract
provisions to allow contractors and grantees the
necessary flexibility with regard to distribution of
reports and other dissemination strategies.

Recommendation D-15. The Department of Education should
give attention to the identification of "right-to-know"
user audiences and develop strategies to meet their
information needs. (Chapter 4)

Perhaps the most neglected audience for evaluation
studies consists of program beneficiaries and their
representatives. We believe that this neglect is not so
much intentional as it is produced by the very real
difficulties of defining this set of audiences in a
reasonable way. In order to more closely approximate the
ideal that all those having a recognized interest in a
program should have reasonable access to evaluation
results, the Department should consider dissemination of
evaluation reports freely to groups and organizations
that claim to represent major classes of beneficiaries of
education programs. Positive, active dissemination to
such right-to-know groups may include such specific
activities as ascertaining their information needs prior
to evaluation design and during the evaluation, preparing
standard lists of groups and organizations to whom
evaluation results are routinely disseminated, and
seeking out comments and critiques of evaluation
reports. Since it is to be expected that such
right-to-know groups will be different for different
evaluations, careful consideration of the appropriate
right-to-know groups should be part of the dissemination
plans that contractors are asked to prepare as part of
their response to RFPs and grant announcements.

On General Management

Recommendation D-16. The Department of Education should
clearly·spell out minimum requirements for monitoring and
compliance reporting and set standards for meeting the
requirements. (Chapter 5)

Such data items as distribution of funds, number and
types of beneficiaries being served, and specific program
services should be defined by the Department so that
local and state agencies will know exactly what reporting
is required of them. Quality control procedures should
be enforced so that adequate performance reports can be
made to Congress. Before setting the requirements,
however, the Department needs to examine its own capacity
to deal with local and state reports in order to avoid
collecting information that is never used because of the
sheer inability of federal staff to deal with the volume
of reports. The objective of this recommendation is to
improve the quality of data needed for accountability
without increasing the burden of response on local and
state agencies. To accomplish both ends, admittedly
somewhat difficult to reconcile, the Department should
consider appropriate development research on what kinds
of procedures would minimize response burden and at the
same time ensure sufficient data quality.

Recommendation D-17. The Department of Education should
examine staff deployment and should establish training
opportunities for federal staff responsible for
evaluation activities or for implementation of evaluation
findings. (Chapter 5)

The Department should consider alternative ways of
using the technical staff within the central unit and the
evaluation staff in other units. The greater the degree
of government involvement in an activity, the greater the
skills and the greater the number of personnel required:
grants and consultancies entail the least involvement,
contracts and evaluation teams configured of government
staff and outside experts more, and in-house studies the
most. The Department should examine the number and types
of positions assigned in light of responsibilities and
workload. It should also examine the academic and
experience background of personnel charged with
evaluation responsibilities. Such personnel should be

22

well grounded in the theory and methodology of relevant
social science disciplines; they should be aware of the
perspectives of the various parties at interest; and they
should have practical program knowledge. Suitable
training programs should be made available to prepare
staff members adequately for their tasks.

Recommendation D-18. The Department of Education should
take steps to simplify procedures for procuring
evaluation studies, carrying them out, and disseminating
their findings. (Chapter 5)

   The Committee is aware that our recommendations for
opening up the system and for involving minority groups
and other parties at interest during various phases will
complicate and prolong the evaluation process. However,
we firmly believe that this can be more than compensated
for by simplifying and improving internal management
procedures now used by the Department.
   The procurement process has become not only
restrictive and inflexible but very costly in internal
staff time and to proposers, though the cost to proposers
is recouped eventually through overhead and in other
ways, so that the government bears the double burden.
Other sources of delay, once a contract or grant for a
study has been awarded, must also be identified and
addressed. This applies particularly to clearance
procedures and to monitor and agency handling of requests
for changes in study design, sampling procedures,
testing, analysis, time frame, and the like. The
Department should consider sanctions and incentives to
encourage timely performance, and it should hold itself
responsible for timely dissemination.
   Our call for timely performance on studies that are
intended to feed into a specific legislative or
management decision in no way invalidates the need for a
more deliberative approach in certain cases. There are
times, especially when an effort is being made to remedy
a problem that is little understood, when it is more
important to promote a variety of studies that explore
emerging leads than to mount a formal study designed to
provide a definitive answer by a specified date. Even in
such cases, however, the pace should be set by the
research process and concerns for its quality rather than
by overly cumbersome management procedures.

34

1
Introduction

In the broadest sense, evaluation has always been done.
In its more narrow modern usage, "evaluation" has come to
mean the use of recently developed research tools and
concepts of the social sciences to develop evaluation
knowledge. What has social-science-based evaluation
contributed to education? Two examples, one of national
scope, the other local, illustrate how such evaluations
illuminate and sometimes contradict judgments derived in
other ways; they thus increase knowledge about what
affects the educational process and how it in turn may
affect educational and social goals.

In 1959 James B. Conant published his widely read
report on the American high school, recommending, among
other things, the consolidation of small high schools
into large comprehensive schools and an increased
emphasis on English composition, mathematics, and
science. His report, based on visits to several dozen
high schools, was essentially the application of his
judgment as an experienced educator to what he saw as
typical practice in better schools in comparison with
less adequate schools. He concluded that, in the better
schools, students were learning more because the
curriculum offered to them was better, there was a wider
variety of courses, teachers were better, facilities were
better, the counseling was better, and so on through a
list of characteristics generally associated with
comprehensive high schools. Hence, Conant concluded that
such schools contributed to the learning achieved by high
school students. Whatever influence Conant's report had

23

on American education, it was certainly widely read and
discussed at the time. Undoubtedly, the report hastened
the process of school district consolidation that was
already under way and helped the emphasis on academic
achievement that was also aided by the Sputnik
accomplishments of the Russians during the same era.

In a broad sense of the word, Conant's volume
constituted an evaluation of our school system; however,
it was not an evaluation in the sense used in this report
because the means by which Conant came to his
recommendations were not based on the concepts and tools
of social science. He generalized what he found to all
schools even though there was no evidence that the
schools he studied fairly represented all American high
schools. Nor did he collect information on the schools
and students in a sufficiently structured way to allow
replication by other observers. In short, Conant and his
colleagues did not follow the procedures of ethnography,
sample surveys, or experimenters; the procedures used
were essentially those of high-level journalism. But,
most important of all, Conant's observations were not
social science because he did not consider alternative
explanations for differences in quality among the more
than 100 schools that he and his collaborators visited.
Were his "better schools" better because of their
curricula, staff, and amount of per-capita student
support, or were they better for some other reason?

In contrast, the later work by James S. Coleman and
his associates (1966) is clearly an evaluation in the
social science sense. His sample of 469 high schools and
959 feeder elementary and junior high schools was chosen
by probability methods to represent fairly the (then)
21,000 high schools in the United States. Achievement
tests were used to measure the learning of large samples
of thousands of students selected from various grade
levels within the sample schools. In addition,
principals and teachers were queried about their own
professional preparation and about the relevant
facilities available within each school, such as library
size, physical education facilities, and age and size of
buildings.

While there were clearly some high schools that
appeared to be fostering higher levels of academic
achievement among their students, Coleman also considered
alternative explanations for school differences, among
which the most important were family background and
community differences among students. His analysis

showed that characteristics of schools, teachers, or
principals counted very little in comparison with family
background. Indeed, the major difference between schools
was accounted for by the differences in the mixes of
students from various backgrounds, with school facilities
and financial expenditures also counting for very
little. This finding profoundly shocked the field of
education. The main policy implication of the finding
was that changing the academic achievement of children
through changing the schools was not going to be an easy
job entailing merely changes in curricula, upgrading of
teachers, or providing more financial support to the
schools.

The importance of testing alternative explanations is
shown as dramatically in a recent study (Robertson 1980)
of the effect of dropping driver education from the
curricula of some Connecticut high schools. In 1976, the
Connecticut state legislature decided to discontinue
subsidizing driver education in the state's high
schools. In response, some of the high schools dropped
driver education entirely from the curriculum while some
retained it, financing the classes from local funds.
Robertson tested the impact of this change on automobile
accidents involving young persons aged 16 and 17 by
comparing the number of accidents in counties in which
driver education was retained with counties in which it
had been dropped. He noted that over a 2-year period,
the number of accidents involving persons aged 16 and 17
declined drastically in the communities that had dropped
the course.

It would have been easy to conclude that driver
education was not efficacious in training careful
drivers, or even that it produced more reckless drivers,
but Robertson tested a number of reasonable alternative
explanations. The most plausible of these alternatives
was indicated by a drop in the number of drivers aged 16
and 17 in those communities that dropped driver
education. In short, in communities in which driver
education was part of the curriculum, young people
received their driver's licenses at an earlier age and
hence, there were simply more people aged 16 and 17 who
drove. If driver education courses do not lead to a
reduction in the number of accidents for 16- and
17-year-olds, it is not because they are not
educationally effective (we cannot draw conclusions about
this one way or another from Connecticut's natural
experiment), but because they encourage more people of
that age to get licenses.

Since Coleman's landmark work, demand for evaluation
has been increasing, in part because the last 15 years
have seen a burgeoning of public programs funded and
managed through the federal government. The intent of
such programs has been to alleviate a wide variety of
societal problems, from unemployment to low reading
scores of some children in public schools, from
substandard housing to recidivism of felons, from drug
addiction to the inadequacies and inequities of the
health care system. But as a number of the programs
failed to live up to the expectations that accompanied
their creation, even as their costs escalated, questions
were raised as to the reasons for the disappointing
performance. In response, federal agencies have
sponsored and conducted a diversity of evaluation
activities, obligating nearly a quarter of a billion
dollars for that purpose in fiscal 1977 and investing
more than 2,000 staff years on the part of permanent
federal evaluation staff (Office of Management and Budget
1977).

Nowhere has the growth of programs accompanied by the
growth of evaluation been more pronounced than in the
field of education. The federal part of public school
income grew from 4.3 percent in 1962 to 8.5 percent in
1974, from $1.6 billion to $6.6 billion (in constant
1977-78 dollars). The most rapid increase came in the
mid-1960s; by 1966 the federal contribution stood at 7.9
percent, close to the current level (Dearman and Plisko
1979). The increase was largely the result of the
landmark Elementary and Secondary Education Act (ESEA) of
1965 (reauthorized and added to several times since, most
recently in 1978), which mandated a number of federally
funded programs to improve the school performance of
disadvantaged children. Title I, which supports
compensatory education for poor children, was, and
continues to be, the keystone program of this
legislation. To date, more than $26 billion in federal
funds has gone to state agencies and local school systems
under Title I (Kirst and Jung 1980).

Evaluation activities lagged a few years behind,
though the first legislative requirement for evaluation
was built into the original Title I legislation. By the
time the program was 7 years old, more than $50 million
had been spent to evaluate it (McLaughlin 1975). Current
federal investment in evaluation of education programs
totals some $40 million a year (see Appendix A), not
including federal funds spent for evaluation at the state

and local levels. The objectives of the evaluations have
been to establish whether programs are in conformance
with legislative provisions, whether programs are managed
effectively, and whether programs are achieving the
desired goals. It was assumed that evaluation would
answer those questions and, moreover, provide information
that could be used to remedy identified deficiencies.

But achieving evaluations that yield answers has been
as elusive as achieving successful programs. Early
evaluations faced technical problems and failed to
anticipate the highly politicised context that surrounded
the programs being evaluated. As evaluators learned to
cope with some of the early problems, more evaluations
were funded, and in 1970 the Office of Education (OE)
established a central evaluation unit (see Appendix A)
and placed at its head an evaluator of some stature. But
criticism has not abated. Those who sponsor evaluations
or are in a position to use them continue to voice their
disappointment, often finding results irrelevant or not
delivered in time for making decisions on programs.

Because of the theoretical and technical problems and
because of questions on its contribution to formulating
social policy, the field of evaluation has been marked by
a considerable amount of self-inspection. A large number
of studies and books have been devoted to analysing
evaluation, gauging its effectiveness with respect to
making policy decisions, developing improved methodology,
and appraising the quality of individual studies. For
example, a recent review of program evaluations (Boruch
and Cordray 1980) cites more than 150 references devoted
to critiques and analyses of individual studies or of the
field in general; another recent comprehensive overview
(Cronbach et al. 1980) cites nearly 200 such references.
And both these works concentrate largely on the field of
evaluation in education.

Many of the published articles and books include
recommendations for improving evaluations and making them
more effective. Yet as the field has grown and consumed
a more visible share of resources, the number of
questions on the quality and utility of evaluations has
increased. The latest expression of dissatisfaction came
from the Congress in 1978 with the reauthorization of
ESEA (P.L. 95-561): it was a congressional demand for
improvement in the methods, integrity, and uses of
evaluations, which led the Office of Education to
commission the present review of its evaluation
activities by the Committee on Program Evaluation in
Education.

AUDIENCES FOR THE REPORT

A critical issue for the Committee was to define the
audiences for its report. We identified the major
audiences as members of Congress and their staff and the
senior executives within the new Department of Education
for two reasons. First, these two groups had made
specific complaints about the effectiveness of program
evaluation and had asked for recommendations on
improvement. Second, most of the literature assessing
the field of evaluation is addressed to its
practitioners, rather than to the sponsors and potential
users of evaluations. In the Committee's view, the
critical self-inspection that has characterized the
evaluation field has been a mainspring of the development
of this rather young branch of applied social science.
While such criticism must continue to provide correctives
to deficient theory and practice (and to be effective,
must speak to its own specialist audiences), it will
continue to miss the mark for those outside the circle of
"experts"--the very individuals and groups who make
decisions about social programs and who are in a position
to commission and use evaluations. This report is
primarily addressed to them, and our recommendations are
for the legislators and the agency executives who seek to
obtain greater effectiveness and use from investment in
program evaluation in education.

In addition to our main audiences, we believe the
report will also be of interest to several other
audiences. One such audience includes state and local
education authorities, who carry out evaluation
activities with federal education funds. In some
instances, our recommendations concern them directly; but
even when this is not the case, they have a stake in how
evaluations are commissioned and carried out at the
federal level because the programs being evaluated are
the responsibility of state and local agencies. Groups
concerned with assuring that federal education programs
meet the goals intended by the legislation are another
audience. An improved evaluation system will provide
information to carry out their oversight function more
effectively. In particular, such information is critical
to groups interested in furthering equal educational
opportunity, the goal of most federal education programs
and mandates. Lastly, though we have made no effort to
address problems from their particular perspective,
researchers involved in carrying out evaluations are an

audience for our recommendations since we intend those
recommendations to have an impact on how evaluation is
done and used.

SCOPE OF THE REPORT

Among researchers, the term "program evaluation"
traditionally has been applied to the assessment of the
impact of a given program. Generally, this has included
answering two kinds of questions: To what degree have
the changes intended by the program been achieved? To
what extent can the observed changes be attributed to the
program? Early in the Committee's proceedings, however,
it became clear that this definition was too limited for
our task and for the audiences of this report. In the
pragmatic environment in which questions are framed about
federal education programs, distinctions between outcome
evaluations--those concerned with the above
questions--and other types of assessment are frequently
irrelevant. Congress and Department officials need to
know how funds are allocated, what kinds of program
services are being delivered to whom, how management of a
program could be improved, what program alternatives are
most effective, and which programs are most
cost-efficient. In developing new programs or changing
existing ones, questions must be answered about the
nature and extent of the need to be met and about the
effectiveness of proposed programs to meet that need. A
considerable proportion of the funds allocated to
evaluation of federal education programs goes to answer
such questions, and even studies concerned mainly with
program outcome include activities (and money) devoted to
those other issues. From discussions with congressional
and Departmental staff, it was evident that the
dissatisfaction with evaluation encompasses perceived
shortcomings in all areas and that focusing only on
program evaluation as defined by the research community
would not address the concerns of policy makers.
Therefore, the Committee has chosen to be inclusive with
respect to the domain of its inquiry. The terms
"evaluation activities" and "evaluation," as used in this
report, cover work undertaken to answer any type of
assessment or planning question having to do with the
allocation of benefits, the nature of services, the
outcomes, or the management of an established or proposed
program. But we have not given equal attention to each

type of evaluation activity; we have concentrated on
those activities for which the methods of applied
research can make the greatest contribution to policy
formulation.

While the Committee has used an inclusive definition
of evaluation, it has concentrated its attention on a
limited number of issues, namely those of greatest
interest to the primary audiences. Congressional concern
with uniform methods and measures is addressed in Chapter
2 in the context of delineating different types of
evaluation procedures and their appropriate use. Issues
of integrity and independence are treated as part of the
discussion in Chapter 3 of how the quality of evaluations
can be improved. Follow-up on evaluations, the third
issue stated explicitly in the congressional request that
led to our study, is subsumed under the more general
topic of the use of evaluation results, which is
considered in Chapter 4. Finally, Chapter 5 responds to
the specific request made by Department officials to
provide recommendations on the organization and
management of evaluations funded with federal education
funds. The recommendations and suggestions in Chapter 5
also take account of implications for management and
organization that derive from the discussions in the
preceding chapters of evaluation procedures, evaluation
quality, and the use of evaluation results.

The report documents some of the ways in which the
evaluation system in education currently operates and the
incentive structure implicit in its operation. The
Committee makes a number of recommendations that, in our
view, would improve the current system. We suspect that
the effective implementation of the recommendations will
have to take into account the incentives of legislators
and upper-level managers in the Department as well as
those of lower-level managers, contractors, and potential
and actual beneficiaries. Time did not permit a thorough
examination of how incentives might be restructured;
instead, we have largely focused on recommendations that
appear feasible within the present incentive system and
that we think can produce improvements in the quality and
usefulness of evaluations.

Some issues that are the subject of much debate within
the evaluation community have been given only passing
attention in the report, such as: the choice between
quantitative and qualitative methods; the relationships
between those who sponsor evaluations, those who carry
them out, and those directly involved with the programs

being evaluated; and a number of technical matters
relating to effective collection of data and appropriate
analytical strategies. Deemphasis of such topics was not
just a matter of lack of time; it reflects the
Committee's view that those topics are less important to
our main audiences and that (particularly in the case of
technical issues) the Committee would find little new to
add to the extensive literature in the field.

Four additional issues pervaded the discussions of the
Committee, though they had not been identified
specifically in the 1978 legislative provision calling
for the assessment of OE's evaluation activities, by
legislative staff interviewed, or by Department
officials. Of these, the most important surfaced during
the very first meeting, namely, how well evaluation
activities address the broad federal mission of equal
educational opportunity. To do so effectively requires
the active participation in the whole evaluation process
of minorities and other groups intended to benefit from
federal education programs—from the planning and design
of evaluations to their ultimate use. The inadequate
consideration of the needs and viewpoints of the groups
intended to benefit from programs affects the kinds of
questions asked about programs, and insufficient
information about the results of evaluations prevents
such groups from knowing how to make programs more
effective.

The second issue developed as the Committee pursued
its questions about the current process of commissioning
and carrying out evaluations in education. As a result
of external regulations and constraints and internal
procedures, the process operates so as to limit severely
the flow of ideas and creativity that must be part of any
effective research effort, including applied research
such as program evaluation. The conditions that have led
to this undesirable state admit of no easy remedy, but
measures must be taken to open up the process if good
evaluations are to be carried out. Opening up the
process is also necessary in order to have greater
involvement by minority researchers and organizations.

A third issue also bears on quality and equal
opportunity, namely, the training of individuals involved
in evaluation, either as performers or as users. The
Committee is not advocating an expansion of the field of
evaluation, but we are concerned that federal mandates
for evaluation generated both by Congress and by the
Department (and its predecessor) have forced individuals

with inadequate preparation into evaluation, particularly
at the state and local levels. One remedy is to
reexamine current evaluation requirements and reduce them
where they are not warranted; a second is to provide
training and technical assistance as necessary.
Opportunities for training can also deal with the
purported shortage of minority researchers and remedy
specific shortcomings among federal evaluation and
program staff.

   A fourth issue became evident as the Committee
reviewed the major themes and recommendations of the
report. Unless the limitations of evaluation are clearly
recognized, disappointment will continue. Ideally,
evaluators are objective and accurate reporters who can
provide and interpret detailed information about a
program. In reality, they may be asked to act as judges
or as support personnel, or they may be perceived as a
necessary but unwelcome program disturbance. As judges,
the verdicts of evaluators may be considered uninformed
by program managers and clients when the evaluators come
from outside the program and biased when they come from
inside. As support personnel, their findings and advice
may conflict with accepted assumptions, policies, and
procedures. As researchers, the constraints on
resources, on freedom to design evaluations, and on
access to information may sharply limit their ability to
investigate some critical questions. Evaluators often
must negotiate with various parties at interest--the
evaluation sponsors; the federal, state, and local
program managers; teachers and principals; parents and
students--providing some service of value to each in
exchange for resources (a program manager's time, a
sponsor's money) and cooperation. And even when an
evaluation has proceeded successfully, the results must
enter a communication stream that contains many other
messages. Evaluation does not and cannot eliminate the
need to manage controversy; at best, evaluators and their
work serve to produce knowledge that can inform decisions
about programs, decisions that must continue to be made
through political and managerial processes.

   Though the report is organized into chapters according
to the topics of greatest concern to our two main
audiences, the four issues of equal opportunity, opening
up the process, training, and the role of evaluation are
woven throughout the text of the chapters. We believe
that addressing the first two of these issues is
indispensable to increasing the effectiveness and

quality, as well as the uses of evaluations;
recommendations relevant to these issues are made in
several chapters. Recommendations on training and
technical assistance appear in the two chapters dealing
with the quality of evaluations and the organization and
management of evaluation activities. As to the fourth
issue, we hope we have been sufficiently sensitive
throughout our work to both the importance and the
limitations of the evaluator's role, even though
constraints of time and space have precluded the full
discussion that this issue deserves.

A question that surfaced several times during the
Committee's deliberations concerned the appropriate size
of the federal investment in evaluation relative to the
federal investment in education programs themselves.
Depending on what activities are included as evaluation,
some 0.3 to 0.7 percent of total federal education funds
are currently spent on evaluation. Several individual
programs have legislatively established ceilings for
evaluation activities sponsored at the national level
(0.5 percent of program funds for ESEA Title I, 1 percent
for Emergency School Aid Act programs), and there are
provisions for the funding of state and local evaluations
within some mandated set-asides for administrative
expenditures. For large programs, a 0.5 percent
set-aside for evaluation will yield a sizable pool of
funds if invested at the national level, but it may be
inadequate if parceled out at the individual school
system level; for smaller programs, it may be reasonable
to spend as much as 10 percent of total program funds
(see Appendix C). Limited questions about accountability
can be answered relatively inexpensively, but to try to
answer complex questions with inadequately funded studies
may turn out to be a waste of resources. The Committee
considered current funding provisions and spending
patterns and makes some recommendations regarding them,
specifically that evaluation funding be separated from
administrative costs and that complex and costly
evaluations not be undertaken without adequate
resources. But we do not see it as our role to determine
the proper size of the total pool of funds to be devoted
to evaluation. The allocation of resources between
programs and their evaluation depends on the importance
assigned to the flow of program funds to beneficiaries
compared to the importance of gaining knowledge about the
programs and accounting for their effects. This
determination is largely a matter of political judgment.

Instead of attempting to determine whether the current
level of spending on evaluation is too much, too little,
or just right, the Committee has focused on how those
funds that are allocated to evaluation can be spent more
effectively and yield more useful results.

2
Defining Evaluation

THE ROLE OF EVALUATION

The literal meaning of the verb "to evaluate" is to
estimate the value of some object or activity. As
applied to education programs, evaluation includes the
set of activities that are aimed at finding out how
valuable a program may be. Relevant questions include:
How serious is the condition that the program is designed
to ameliorate? How is the program supposed to work?
What would happen without the program? What would happen
if the program were expanded? How valuable is the
program compared to other programs?
    Putting things this way makes it very difficult to
question the value of evaluation. How can one be for not
knowing the value of a program, its impact on this or
that, or what would happen if it were altered? How can
one favor making budgetary decisions in the absence of
evaluation information of some sort? In short, how can
one opt for ignorance over knowledge?
    Although the need to know seems indisputable,
controversy and struggles inevitably arise whenever
social-science-based evaluations are done and reported.
First, such evaluations make program goals explicit and
thus may uncover previously hidden value disagreements.
Second, they have to compete with other forms of
evaluation--ad hoc opinions, skillful journalistic
reporting, intuitive perceptions, and so on. Third, the
evaluation process is rarely clear cut or simple: a
given program can be evaluated using a variety of
alternative research methods, and results are often
subject to competing interpretations. For these reasons,

evaluation through social science methods tends to be politicized: it cannot help but be influenced by political tides, varying ideological perspectives, personal goals and inhibitions, technical limitations of methods used, economic priorities, etc.

A special difficulty for evaluation is the fact that scarcely anyone likes to be judged, and those who run and operate programs or benefit by them are especially likely to react defensively to such judging. Even if the results of the evaluation may be favorable, the scrutiny is difficult to tolerate. There is always the concern that one's behaviors, attitudes, and beliefs will be misinterpreted and distorted in a professional language that is incomprehensible or presented in a form that robs one's individual identity. But beyond the personal concern that one will be misunderstood or misinterpreted is the recognition that evaluations necessarily represent some particular point of view and reflect specific value positions. By their very nature, evaluations are not neutral. Judgments are made based on implicit or explicit assumptions about what a program is and what it should be. To those running a program or benefiting by it, evaluators' judgments are often considered external to the program and hence inappropriate.

It is obviously important that evaluations be undertaken by persons who are not deeply committed to or involved with the program being evaluated because their special interests and deep connections are likely to blind them from seeing the program's inadequacies and weaknesses. But it is also true that the distance and dispassion of an external observer do not necessarily lead to objectivity. Distance and dispassion can also lead to disengagement from what is going on, a lack of identification with and empathy for those who deliver program services and those who receive them, or even worse, an alienation from and disregard for the objectives and values held by them. Good evaluators must balance precariously between an intimate and responsible knowledge of the program and a distance from it that will permit them to see its strengths and weaknesses.

The evaluation process is further complicated by having many diverse audiences that may be eager to know about the impact and effects of the programs being evaluated. Each audience tends to have its own needs for and expectations about information. With various agendas and levels of sophistication, such diverse audiences make a variety of demands on evaluators, sometimes contradictory ones.

For education programs, Congress and the Department of
Education constitute two highly visible and crucial
audiences. They are crucial for two reasons: first,
they can make the decisions about which program to
initiate or to expand, which to discontinue or to
contract; second, they fund evaluations. Although the
scope and responsibilities of the Congress and the
Department of Education are clearly the broadest, they
are not the only audiences to whom evaluators of
education programs must address their findings. Program
decisions about education in the United States (even of
federally supported education programs) are only partly
made at the federal level: thousands of school boards in
local communities make most of the school policy that
affects the specific character of public education.
State education agencies (SEAs) also affect what is
taught and how it is taught in each of the 50 states.
These local and state school authorities may be able to
use information provided by evaluations if the findings
are presented in ways that are relevant and
understandable. Indeed, not enough careful thought and
attention has been given to the problem of how such
information can be provided in the most understandable
and relevant ways.

Perhaps the greatest impact of evaluations is on those
who manage education programs and those who provide the
services of the programs. They are the people whose work
is being judged. These audiences have the most direct
involvement in the programs, are most likely to be
threatened by the evaluation process, and may be very
fearful that programs will be curtailed or cut off
because of an evaluation's findings. Program personnel
are, understandably, usually more concerned with the
protection of their own programs and projects than they
are with the advancement of knowledge. Their political
power can be and has been exercised to save a program
that appears to be threatened (for example, Head Start,
Impact Aid). Often, a negative evaluation finding for a
national program appears unjust to local program
personnel, who believe that their projects may be better
than the average, and offers little help to committed
staff who wish to make improvements. Nevertheless, some
forms of knowledge from evaluation can be useful to
program personnel, to teachers and administrators, for
example, who want practice-oriented information that may
help them provide more effective instruction.

The consumers of the services provided by a program--parents and their children--also have a stake in evaluation, although rarely have national evaluations been addressed to this audience. This audience is often the most elusive of all because it is not always articulate or well organized. When the consumer audience has been organized, it has usually been in favor of saving a program despite appparently negative findings, probably in the belief that it is better to have a program, even if its effects cannot be proved, than to have no program at all. Yet it is not clear whether consumers have more of a stake in the continuation of a program, regardless of its success, or in the continual improvement of education through development and evaluation of program alternatives. We believe that the consumers of education programs have been the most neglected of all potential audiences; although we recognize that to develop this potential audience into an actual one will require much experimentation with alternative modes of communication.

To further complicate the picture, there are other overlapping constituencies and special interest groups that are concerned about evaluation processes and findings. These groups often reflect minority perspectives that they feel have been neglected or ignored by traditional evaluation designs and outcome measures. They argue for the inclusion of their perspectives in the goals, methods, analyses, findings, and recommendations of evaluations. The National Urban League, for example, which has its own sophisticated research department, has been interested in the evaluations of special programs designed to increase the reading scores of inner-city, minority children. It carefully monitors the programs (value assumptions as well as instructional methods) as well as the evaluation strategies, the data, and the language and style in which findings are presented. The National Organization of Women and other feminist groups carry out similar monitoring of programs and of related evaluations that are of concern to them. These special interest groups are becoming increasingly visible audiences, and they seek to intervene at various points in the evaluation process.

In some sense, an evaluator is expected to provide feedback to all of these audiences, an often baffling and unrealistic expectation, for each of them has a different kind of stake in evaluation, speaks a different language,

and has a different conception of usable knowledge. This
report argues that all of these audiences are important,
but that any particular evaluation usually should not try
to be responsive to all of them. Responding to the
myriad and often conflicting expectations of all the
audiences is likely to diminish the integrity of an
evaluaton and limit its usefulness to any one audience.
  The "primary" audience(s) of an evaluation should be
identified by those who call for it and by the evaluators
who carry it out: the design of an evaluation should
anticipate the primary audience(s), and the procedures,
methods, analysis, and the language of its reports should
correspond to the needs and expectations of the primary
audience(s). This does not mean that the findings of an
evaluation will be useless or wholly irrelevant to the
"secondary" audiences, but it is likely that there will
have to be some amount of translation and
reinterpretation to make the information useful to them.
Defining the audience and targeting the message will
reduce the frustration that often accompanies the more
eclectic attempts to speak simultaneously with many
tongues to many groups. Inevitably the selection of the
primary audience(s) becomes a controversial process, one
that must be endured, coped with, and responded to by the
evaluator. In the case of evaluations that are mandated
by Congress or commissioned by the Department, the
mandate should include some designation of the primary
audience(s) to which the evaluation is addressed, as a
guide to the evaluators.
  The evaluation process is necessarily a controversial
one that requires more than technical and procedural
solutions. Technical matters and procedures are not
unimportant, but there are other important demands that
must be managed with equal care. Those demands include
resolving the tensions among opposing values and
perspectives, dealing with political priorities, and
taking account of contrasting methodological traditions.
Most of this report focuses on evaluation strategies and
objectives, issues of quality control, utilization of
findings, and the organization of evaluation structures.
Although these technical and substantive questions are
critical to those seeking to improve evaluation studies
in education and increase their usefulness, it is
important that the evaluation process be seen in context
and that the reader be cognizant of the myriad forces
that combine to shape any evaluation.
  Evaluators must respond to these contextual issues:

51

they must play a role that includes being aware of the
primary and secondary audiences and of competing
constituencies, finding the appropriate distance from the
programs that will permit access and understanding but
not lead to distortion, and seeking to neutralize their
place in a highly political environment. Although the
role of an evaluator is in many respects a responsive
one, it should not be viewed as essentially reactive.
Evaluators must do more than negotiate among competing
interest groups or respond to the various priorities and
needs for information. Unless they maintain some measure
of autonomy, they will be useless to all those who call
on their services. It is critical to be aware of the
needs of the various interest groups, but a keen
understanding of audience perspectives should not mold
the entire shape of any study. In moving beyond the
reactive mode, evaluators might well be envisioned as the
translators and bridge builders among the various spheres
of research, policy, and practice. Because their work
requires that they be adaptive to several environments,
they have a unique opportunity to find ways of
translating and interpreting knowledge and understandings
from one environment to the other.

THE VARIETIES OF EVALUATION

A decade ago, social scientists carrying out evaluations
tended to concentrate on providing estimates of the
relative effectiveness of programs. As experience
accumulated, however, it became increasingly clear that
more knowledge was also needed in designing, improving,
and implementing programs. Hence, the scope of
evaluation has been enlarged to include research in
support of policy formulation and program development.
The diversity of research activities being carried out
under the general term "evaluation" has led to some
misunderstandings, especially between evaluators and
policy makers. On occasion, policy makers have used
"evaluation" to mean research of a particular sort, while
evaluators have interpreted "evaluation" to mean a
completely different type of research.
    In an effort to improve the terminology employed in
evaluation activities and to make the terms used more
specific in their meanings, we outline in Figure 1 the
various uses of social science research in support of the
design, implementation, and assessment of social

Questions Arising During the Formation of Policy and the Design of Programs

Policy Question	Evaluation/Social Research Procedure	Research Methods Used
A. How big is the problem and where is it located?	Needs assessment	Assembly of archived data (Census, NCES, etc.) Special sample surveys Ethnographic studies
B. Can we do anything about the problem?	Basic research	Assembly of archived research studies Specially commissioned research
C. Will a proposed program work under optimal conditions?	Small-scale testing	Randomized controlled experiments Pilot studies and demonstrations
D. Can a program be made to work in the field?	Field evaluation	Ethnographic studies Randomized experiments Field tests and demonstrations
E. Will a proposed program be efficient?	Policy analysis	Simulation Prospective cost effectiveness studies Prospective cost-benefit analyses

Questions Arising for Enacted and Implemented Programs

Policy Question	Evaluation/Social Research Procedure	Research Methods Used
A. Are funds being used properly?	Fiscal accountability	Fiscal records Auditing and accounting studies
B. Is the program reaching the beneficiaries?	Coverage accountability	Administrative records Beneficiary studies Sample surveys
C. Is the program implemented as intended?	Implementation accountability	Administrative records Special surveys of programs Ethnographic studies
D. Is the program effective?	Impact assessment	Randomized experiments Statistical modelling Time series studies
E. Is the program efficient?	Economic analyses	Cost effectiveness studies Cost-benefit analyses

FIGURE 1   Policy questions and corresponding evaluation procedures.

programs. The remainder of this report draws upon the
terminology established in Figure 1. Both the figure and
the discussion below project a degree of linearity
associated with policy formulation and program management
that is obviously at odds with reality: programs are
more frequently than not enacted before systematic needs
assessment and program testing have taken place; after a
program is implemented, some monitoring questions are
asked too early, others not at all; changes are made in a
program before there is evidence about it, let alone
evidence on the likely effects of the changes. Our
discussion of the different types of evaluation questions
as applied to education programs is sequential in order
to simplify mapping the terrain, not to indicate the
order usually followed--or necessarily appropriate in
every instance.

Evaluations for Planning Programs

We draw a basic distinction between evaluation questions
that arise during the planning of programs and those that
arise after a program is operating. The first half of
Figure 1 shows the evaluation questions that usually
arise during the planning of a program, along with the
social science research procedures that are generally
employed to provide answers to those questions.

Needs Assessment

Logically, the first question shown in Figure 1 should be
asked at the outset of discussions about policy. An
educational problem has been identified, but questions
may arise about the size of the problem and where it is
concentrated. Thus, illiteracy may be identified as a
problem, but there may be little information on how many
illiterates there are in the nation or whether there are
a disproportionate number among some age groups, ethnic
groups, or regions of the country. The social science
research designed to answer such questions has come to be
called needs assessment.

The research effort involved in providing answers to
the needs assessment question can be as inexpensive as
copying relevant information from published reports from
the U.S. census or as expensive as several years' effort
involving the design, fielding, and analysis of a

large-scale sample survey, such as the study by Coleman et al. (1966) on equal educational opportunity. Needs assessments do not have to be undertaken solely with quantitative techniques. Ethnographic research may also be instructive, especially in getting detailed knowledge of the specific nature of the needs in question; is likely to be especially effective in determining the nature of a need and understanding the processes involved in the generation of a problem. Formal quantitative procedures, however, are essential when the extent of the need has to be established. Obtaining accurate, up-to-date data on the size and distribution of a problem, such as illiteracy, is an important first step in planning. Assessment of need and of the contexts in which the need is prevalent will help define the problem. Needs assessment will also help determine the size of a program and attendant costs, at least in part.

Basic Research--Choice of Intervention

The second question concerns whether anything can be done about the problem, and if so, what intervention appears the most promising. Answers to this question depend largely on how much is understood about the problem and what policy-related factors can be changed to affect it. Basic research is the activity that provides the answers to this question. Hence, long-range support for basic research on educational processes is critical for the development of the fundamental ideas for education programs. For example, it is necessary to know why there is a connection between socioeconomic level and the rate of learning of basic skills by children in order to properly design programs to improve the learning rates among children from the lower socioeconomic levels. It is also necessary to know how much such learning rates could be improved by changing teaching methods, by lengthening the school day, or by any other policy measure that could be translated into a program. Even when the ideas for such interventions come from seemingly successful exemplary practice rather than from fundamental theory, basic research is necessary to establish the causal connections between the interventions and the learning effects in order to identify the critical components that make the practice successful and, hence, replicable.

At the time that one is looking for proposed

interventions to ameliorate an educational problem,
commissioned review papers may be an easy way to bring
together relevant existing findings from basic research
since the diverse technical literature dealing with
educational processes is often difficult to master.
However, basic research often does not address suitable
policy variables because basic research is concerned with
the total causal system as it creates a problem, while
the variables that can be changed by policy may be only a
small part of the system. For example, studies of
children who are disciplinary problems in school may
stress understanding the links between the family
situations of the children and their behavior. But for
policy and programmatic purposes, it would have been
considerably more useful if there were studies of how
disciplinary systems within schools affect the rates at
which disciplinary problems appeared within schools.
General research consciously linked to the role that
schools and the educational system generally play in
learning and other behavior may be the best answer to
policy needs. Such research may take a variety of forms,
ranging all the way from systematic observational studies
of school children to carefully controlled randomized
experiments that systematically vary the policy-relevant
experiences of children. Without slighting basic
research support, it should be emphasized that such
policy-relevant general research needs special grant and
contract research programs with review personnel that are
familiar with what is relevant to policy.

Small-Scale Testing--Program Development

Given a promising intervention, the question that next
arises is whether a specific program design will work.
Pilot testing of proposed programs through experiments
and demonstrations can often lead to better information
on whether and how such programs might work. Thus, the
contract-learning experiments funded by the Office of
Economic Opportunity in the early 1970s showed that,
while some contractors could provide effective learning
experiences, the program aroused considerable opposition
among teachers and school systems and hence would not be
a successful program if the program mandated the use of
outside contractors (Gramlich and Koshel 1975).
    We advocate the use of randomized controlled
experiments at this stage in the development of a program

because they are powerful. But because they are also
expensive, the scale should be relatively modest. The
great virtue of randomized controlled experiments is that
they eliminate the possibility that effects may be caused
by processes other than the intervention; hence, they
give a potentially useful program the most valid test.
Moreover, program administration can be controlled to
ensure that the intervention takes place as intended.
Under such conditions, a program has the maximum chance
of working: if it is not effective when carried out
under controlled conditions by dedicated researchers,
there is no reason to believe that it will work under any
conditions. However, a commitment to randomized
experiments for testing programs should not minimize the
complementary potential of ethnographic studies at this
stage, particularly to document why a particular
intervention succeeds or fails.

Field Evaluation--Program Delivery

Even if small-scale testing demonstrates a program's
effectiveness, it should often be changed before being
widely adopted. The relevant question is how properly to
adapt a proposed program so that it will be effective
when it is no longer under the control of researchers or
specially trained personnel. Unless the program can be
made to work in school systems and in the hands of their
personnel (or other intended service deliverers), it will
not alleviate the problem it is supposed to address, no
matter how effective it was in the experimental setting
(Rossi 1979a). A process of mutual adaptation often
takes place (Berman and McLaughlin 1975-78) that changes
the program as carried out in a given site as much as the
site is changed by the program. Changes that are likely
to be made by the people and institutions that will be
responsible for program delivery must be understood and
built into the program in such a way that effectiveness
is maintained or even enhanced. Field evaluation
(sometimes called formative evaluation) uncovers the ways
in which programs can be changed so that they will work
well within existing educational settings.
Unfortunately, such field testing has not been undertaken
in a systematic way for many education programs, although
it has been done in other social service fields: the
national supported-work demonstration (see Manpower
Demonstration Research Corporation 1979, Maynard et al.

1979) tested a program of transitional, subsidized work experience for people with long-standing employment problems; the youth entitlement demonstration (see Diaz et al. 1980) tested the notion of linking a job guarantee to school attendance and performance.

Randomized controlled experiments are again an extremely powerful tool at this stage; optimally, they should be used to compare several alternative modes of delivery. They should be accompanied by process research activities that use sensitive and observant researchers in close contact with field testing sites. Ethnographic accounts can be extremely useful in understanding why programs do or do not work as anticipated, how the specifics vary from site to site, and what processes impede or facilitate implementation.

Policy Analysis--Program Efficiency

Finally there is the issue of whether a program will be efficient, a question that is answered through prospective policy analysis. Here the issue is how much the program will cost, how much service will be delivered at what level of cost, and whether the anticipated costs of the proposed program overshadow the anticipated benefits. Simulation and prospective analysis, using data from small-scale tests and from field evaluations, are inexpensive and ought to be performed before a program is enacted into law or widely adopted.

Evaluations of Existing Programs

The second half of Figure 1 shows the evaluation questions that arise after a program has been enacted and is in operation.

Fiscal Accountability

Studies of fiscal accountability are perhaps best understood by all since they are part and parcel of the long tradition of auditing the books of public agencies. Procedures are well established and hence much less problematical than those for other types of evaluation activities. In federal education programs, often the only fiscal information comes from grantees' reports on

the use of federal dollars; usually only the large
programs are audited by federal auditors. Fiscal audits
tend to overlap with other forms of evaluation when
questions are also asked about how the money was used
(not just whether it is accounted for). Since
conventional accounting categories are generally not
sufficiently sensitive to determine the level of services
being delivered, the fact that funds appear to be
appropriately spent in an accounting sense does not
necessarily mean that program provisions are being
carried out as intended. Fiscal accounts cannot
establish program integrity, nor can such accounting
establish the true cost of programs, since it does not
consider hidden or opportunity costs.

Coverage Accountability

A significant substantive issue is whether a program is
reaching the population that is intended to receive its
benefits. It should be noted that this issue often turns
out to be of considerable importance: not infrequently,
programs do not reach their intended beneficiaries or
they reach persons who were not intended to be
covered--as was the case for Title VII bilingual
education programs (Danoff 1978) and for the television
program "Sesame Street" (Cook et al. 1975)--or both.
Studies designed to measure coverage are similar in
principle to those discussed under "Needs Assessment"
above. An important source of data for this kind of
evaluation is a program's administrative records, which
often help to identify overcoverage where this is a
problem. Undercoverage, however, may often involve
special surveys.

Implementation Accountability

Questions about how a program is being implemented entail
studying whether and how intended educational services
are being provided. There are many ways in which a
program can be less effective in the field than
expected. Local program personnel may not be properly
instructed in how to administer the program because
school and teaching staff may not have received needed
in-service training. Regulations may be unnecessarily
confusing. The local context may militate against

administering the program as intended, perhaps because
resources presumed to be present may not be. Funds
intended for a particular program may be used to
substitute for funds formerly furnished by other
sources. Programs that require institutions to apply for
grants to extend benefits to the target population may
not be presented in attractive enough terms to achieve
adequate participation rates. As a result, fine-tuning
of basic legislation or of administrative regulations may
be required.

This kind of evaluation is sometimes also labeled
process research, because the questions being asked
concern the nature of a program as it is actually being
delivered and experienced at the particular sites and by
the persons involved there. Such evaluation may be
relatively simple or may involve measurement problems of
considerable complexity. Thus it may be very easy to
learn from schools how many hours per week their new
computer terminals are being used, but very difficult to
learn what precisely is going on inside a classroom when
teachers attempt to use a new teaching method, when
classroom organization is changed, or when other services
are introduced that are highly dependent on persons for
delivery. Studies that require direct observation and
measurement of classroom activity may turn out to be very
expensive to carry out on a large scale. However, for
purposes of fine-tuning a program, it may not be
necessary to proceed on a large scale: it may not matter
whether a particular problem in implementing a program
occurs frequently or infrequently, since if it occurs at
all it is not desirable. Hence, small-scale qualitative
observational studies may be most fruitful.

Impact Assessment

Is a program effective? To answer this question is a
task that requires the highest level of social science
research skills. The essential issue is whether a
program produces more of an intended effect than would
have occurred without the program. While the question
may appear to be simple, impact assessment is extremely
difficult to carry out well. It entails both the
statement of some measurable goals and the determination
of what would have happened without the program. Each
step is difficult. Negative effects must also be looked
for. Even when measurable goals are agreed to and the

60

differences made by the program can be determined,
distinguishing between success and failure is not a
clear-cut decision; there are usually degrees of success
or of failure. A program intended to improve reading
that succeeds in raising students' average reading level
by a half-year more than expected (in the absence of the
program) is less successful than one that has
effectiveness estimates of a full year.[1] This
quantitative difference has to be translated into a
qualitative difference when the decision to fund one
rather than the other program comes into question.

The critical effectiveness issue is whether a program
does anything for its beneficiaries to help them advance
towards the goals of the program. While it is relatively
easy to measure the status of beneficiaries at any time,
the difficult problem is to determine what their status
might have been had they not participated in the
program. An ideal solution to this problem is the
randomized controlled experiment, which ensures that the
people within the experiment who participate in a program
are "identical" to the people in control groups who do
not participate in the program. Randomized controlled
experiments, however, are usually not feasible for
studying programs that have been in operation for some
time, since it is ordinarily not possible to find
appropriate individuals who have not been exposed to the
program to assign to control and experimental groups. As
suggested above, such experiments are most appropriate in
the program development phase. For ongoing programs,
other techniques must be employed, such as comparing
participants before and after a program has been enacted
or comparing beneficiaries to those who do not receive a
program's benefits. Such research and statistical
techniques require extreme care; a large literature that
is devoted to them warns of the many pitfalls in their
use.

Policy makers should call for impact assessment only
when circumstances warrant such studies (see below).
They should be wary of requiring impact assessment from
agencies that cannot marshall the necessary skilled
personnel. They should be equally wary of requiring
impact assessment, which is expensive to do adequately,
without providing sufficient funds. In particular, only
a few local and state education authorities have the
capabilities or resources to competently carry out impact
assessments; hence, such tasks should not be imposed on
all state and local agencies without special attention to
providing sufficient resources.

61

Economic Efficiency

The final question in the second half of Figure 1 asks
whether the costs of ths program are justified by the
gains achieved. The same question might be raised in a
comparative framework, that is, whether program X is more
efficient than program Y in achieving some particular
goal. While these questions also arise during the
planning phase of program development (see above) at
this point in the process the answers are no longer
anticipated costs and benefits but actual costs and
benefits based on good estimates of effectiveness and
field experiences with the programs.

The main problem in answering such questions centers
around establishing a yardstick for such an assessment,
for example, dollars spent for units of achievement
gained, for number of students covered, or for classes or
schools in the program. The simplest way of answering
questions of efficiency is to calculate
cost-effectiveness measures, for example, dollars spent
per unit of output. In the case of the "Sesame Street"
program, several cost-effectiveness measures were
computed, such as dollars spent per child-hour of viewing
and dollars spent per additional letter of the alphabet
learned (Ball and Bogatz 1970, Bogatz and Ball 1971).
(Note that the second measure implies knowing the
effectiveness of the program, as established by an impact
assessment.) The most complicated mode of answering the
efficiency question is to conduct a full-fledged
cost-benefit analysis in which all the costs and benefits
are computed. Relatively few full-fledged cost-benefit
analyses have been made of social programs because it is
difficult to measure all the costs and all the benefits
in the same terms. In principle, it is possible to
convert into dollars all the costs and benefits of a
program; in practice, however, it is rarely possible to
do so without some disagreement on the valuation placed,
say, on learning an additional letter of the alphabet.

WHETHER TO EVALUATE

Implicit in the preceding discussion is the assumption
that a program, prospective or enacted, can be evaluated
in some way or another; however, that is not always
true. There are some programs, whose characteristics are
described below, that cannot be fully evaluated or that
cannot be evaluated at all.

All programs that have been enacted can be evaluated
in the sense of fiscal accountability. Procedures that
have been detailed in laws or in regulations can also be
evaluated as to whether they are being carried out as
intended. But only programs that specify clearly the
intended beneficiaries and the intended effects can be
evaluated fully. This is not to say that programs with
vaguely stated aims are not worthwhile; it is to say that
they cannot be evaluated as to their effectiveness.
Thus, a program that has the announced intention of
enriching the cultural lives of high school students
cannot be evaluated with respect to its impact because
the aim of "enriching the cultural life" is simply not
specific enough to provide criteria for judging
effectiveness. In addition, the group of intended
beneficiaries, high school students, is so broad and
inclusive that one simply could not measure "effects" for
all of them.

A prime requisite for being able to evaluate the
impact of a program is the existence of clearly
designated, specific aims. But, as Wholey et al.
(1975:89) note:

> As a natural result of the political process,
> federal programs usually have many poorly defined
> objectives. Authorizing legislation and program
> guidelines are generally vague about program
> objectives and priorities. . . . Policy-makers and
> managers often perceive that ambiguity about what
> constitutes success is an asset, permitting
> flexibility and helping ensure survival.

This situation often puts evaluators in the position of
setting goals or selecting among several stated goals. A
program may have a number of diverse goals: for example,
Head Start was intended to provide better health care and
nutrition for poor children, improve their cognitive
development, increase their social competence, improve
the conditions of participating families and communities,
serve as a focus for political action and community
organization, and result in more effective functioning of
other service agencies. (See, for example, Office of
Child Development 1973.) In such cases, evaluators and
those who commission evaluations must agree on which of
the goals are most important to assess and whether they
are sufficiently specific to permit an impact
evaluation. Often, however, the problem of goal

selection is governed by the law of instruments: as the early evaluations of Head Start demonstrate, those goals for which measurement instruments exist--for example, cognitive achievement--will be the goals by which a program is evaluated, even though other goals may be equally important.

Some programs allow each local school system to set its own goals within broad program aims and to design its own interventions, provided money and services go to the target population. For such a program, it is possible to evaluate the impact of individual local projects but nearly impossible to gauge the effectiveness of the overall program by aggregating effects over many sites. A similar problem exists for programs that provide funds or other assistance to local school authorities without specifying more than very general goals. These, too, cannot be evaluated for impact at the national level because there is, in fact, no national program but a collection of diverse local programs. For example, Title I of ESEA is intended to expand and improve education programs for educationally deprived children but it does not specify in any detail what is to be accomplished. Therefore, it cannot be evaluated nationally (except in the accounting sense), though projects at individual sites can be evaluated if goals and interventions are sufficiently specific.[2] Indeed, programs like Head Start and Title I have never been successfully evaluated for national impact no matter how massive the study without heroic assumptions concerning their intended aims, assumptions that then created considerable controversy when evaluation findings were released. Results from individual local studies may cumulate as a program matures, however, and should be synthesized to permit general conclusions.

This criterion of specificity in aims also applies to prospective programs. If such programs do not have specific aims, they cannot be developed properly using social science evaluation unless sponsors are content to let evaluators specify program goals and intended outcomes. Experiments and demonstrations cannot be properly designed without knowing what the criteria for effectiveness are to be; cost-benefit analyses cannot be made without knowing what the anticipated benefits are; and so on.

Techniques have been developed (Wholey 1979, Schmidt et al. 1979) to determine whether a program can be evaluated (in the senses discussed above), i.e., whether

it is evaluable. Members of Congress and other decision
makers may want to commission such studies of
evaluability as a first step in evaluation rather than to
assume that all programs can be evaluated. Indeed, we
commend the Department for shifting some of its
evaluation resources in this direction; so far, 10
evaluability studies have been commissioned by the
central evaluation unit of the Department.

WHEN TO EVALUATE

Even if a program is sufficiently specified to allow both
accountability and impact evaluations, conducting impact
evaluations may be inappropriate at a particular time
because of the stage of program development or
implementation. There are three phases in the life of a
program that are notably inappropriate for impact
evaluations. The first is during the program's
development. We have suggested that a proposed program
be tried out under actual field conditions after it has
been proved to be effective in a controlled experimental
setting. The purpose of this phase is to adapt the
program so that it will be maximally effective under
normal operating conditions. Obviously, impact (or
summative) evaluation is totally inappropriate during
this phase; at this point, evaluation should be used as a
tool to fine-tune the program, not to judge it.

The second phase is after a program has been enacted
and is being put into operation. All programs require a
shakedown period, during which program administrators
develop regulations and operational procedures and
teachers and school personnel (or other service
deliverers) become familiar with the program's objectives
and methods. The more complex a program, the greater the
start-up problems. When a program allows flexibility and
local choice, further time must be permitted for local
decision making and development of specific features.
Until a program has stabilized, it ought not to be
evaluted, except for fiscal accountability. Too many
negative findings have, in the past, been due to
premature impact evaluation. Even accountability
evaluations may be inappropriate in the early
implementation stage, as demonstrated by findings on weak
administration and even misuse of Title I funds in the
first studies of the program, findings that did not hold
up once personnel at the state and local levels had

learned how to operate the program (Kirst and Jung
1980). The Title I studies also demonstrate another
point: if more effective policy analysis were conducted
before implementing a program to ensure that program
legislation and regulations did not lead to confusion in
the field, the shakedown period might be considerably
reduced.

The third phase during which impact evaluations are
inappropriate involves education programs that have
long-range as well as short-range objectives. For
example, career education may be concerned with helping
youth achieve both entry-level skills and satisfactory
career paths. Obviously, the second objective is not
measurable until effects emerge after a number of years.
Assessment of such effects requires time-series studies,
which take long-range commitment or sophisticated
statistical modeling that requires highly skilled
researchers. Too often, impact evaluations have either
ignored long-range effects as too costly and
time-consuming to assess, or they have attempted
assessment of long-range effects in an unrealistic time
frame. As a result, the full effects of the program
remain unknown, even though evaluation is said to have
taken place. If programs are to be judged by their
results, enough time must be allowed for the program's
full effects to emerge before full-scale impact
evaluation can be done.

One final point about the timing of evaluations
concerns old programs. There is a need to address policy
issues in programs that have been operating so long as to
become routinized. How have conditions changed? Are
there different educational goals? Have the needs of
intended beneficiaries changed? Periodic evaluations may
provide needed "shake-up" to ensure that a program is
still meeting priority objectives.

Recommendation C-1. When Congress requests evaluations
it should identify the kind of question(s) to be
addressed.

At present, there is a multiplicity of requirements
for evaluation that vary from title to title (see Boruch,
Cordray, and Pion, Ch. 3 in Boruch and Cordray 1980). In
some cases, Congress calls for elaborate and detailed
evaluation studies involving sophisticated quantitative
techniques and analyses; in others, requests are made for

impressionistic and anecdotal reports. Congress needs to
be more systematic in its approach to evaluation.
Instead of specifying methods, Congress should make sure
that evaluators are clear about the questions to be
answered.

Figure 1 above identifies 10 kinds of evaluation
activities. At least part of the charge that evaluations
have been irrelevant to Congress's needs for information
stems from the fact that Congress has often been
interpreted to be calling for impact evaluation when in
fact it desired only to know, say, how well a program was
meeting its coverage requirements. A call for evaluation
that does not specify what questions are being asked can
lead to the mismatching of expectation and performance by
Congress and the evaluators. While legislators might
include the policy questions to be addressed directly in
the legislative provisions for evaluation of a program,
it may not always be possible to frame questions with
sufficient specificity at the time evaluation provisions
are being enacted, especially for new programs. In such
cases, sufficient dialogue should take place between the
legislators and the implementing agency and the
evaluators to ensure that the evaluation will meet its
intended objective (Berryman and Glennan 1980).

Congressional mandates for evaluation should also
identify the audience that is to be served by the
legislated evaluation: Congress beneficiaries such as
parent or other interest groups, local program
administrators, federal program administrators, and the
like. The reasons for specifying audiences in any
evaluation are discussed in greater detail in later
chapters. The reason for including audience
specification in this recommendation is that such
specification will also sharpen the policy questions
because different audiences tend to have different
information needs.

Though we recommend that it be specific with respect
to question and audience, legislative language regarding
evaluation should refrain from specifying details of
method (such as sampling procedure or use of control
groups) or of measurement. These are matters requiring
careful technical consideration of specific evaluation
conditions and contexts and should be chosen only after
adequate planning and the application of expert knowledge.

Recommendation D-1. In evaluations initiated by the
Department of Education, the kinds of evaluation
activities to be carried out should be specified clearly
and should be justified in terms of program development
or program implementation.

   This recommendation is analogous to the one to
Congress, but emphasizes the need to think through what
type of evaluation activity is appropriate at any given
stage of development or implementation of a proposed or
an existing program. While evaluation activities are, of
course, specified in great detail by evaluation personnel
at the procurement stage, this recommendation is directed
to the overall evaluation planning stage when top-level
Department officials need to specify what they wish to
know about a program (i.e., the policy questions), why
they wish to know it at some specified time, and what
other audiences have information needs that must be
satisfied through evaluation activities.

Recommendation D-2. When pilot tests of proposed major
programs are conducted, pilot tests of evaluation
requirements should be conducted simultaneously to
determine their feasibility and appropriateness.

   One of the welcome procedural improvements in recent
years has been the greater use of pilot tests of proposed
national programs. The argument is often made that pilot
tests and field evaluations are costly and time consuming
and that an urgent social need cannot remain unaddressed
while the ponderous process of research proceeds. But
the urge to get programs off the ground without prior
testing brings with it certain and often high costs:
programs develop an array of self-interested suppliers
and clients who are likely to fight any changes, even
when subsequent evaluations and research indicate that
they are needed. The Committee endorses the concept of
pilot tests since they have the obvious advantage of
allowing decisions on implementation and on program
changes to be made before programs become entrenched.
Another welcome precedent is that, more and more,
legislation routinely prescribes that programs contain
their own evaluation requirements. Such provisions
ensure that some sort of evaluation will be made of
programs on a continuing basis.
   This recommendation focuses on the intersection of

these two developments. While pilot tests of a program
are being made, it is relatively easy to also conduct a
pilot test of the proposed evaluation. Such a pilot test
can be used to find out what measurements can and cannot
be made of program benefits, how programs should account
for and measure costs, which testing instruments and
procedures are disruptive and which are not, how large a
sample of beneficiaries is needed to get valid program
measurements, and so forth. If a pilot test of the
evaluation is carried out in conjunction with the pilot
test of the program, the design of both the program and
of the evaluation requirements will be strengthened.
Indeed, if evaluation requirements are not pilot tested,
it is difficult to see how those charged with evaluation
responsibilities at the local and state levels are to be
held accountable.

STANDARDIZATION OF METHODS AND MEASURES

As indicated in the preface to this report, one of the
missions given to the Committee was to make
recommendations and proposals ". . . to ensure that
evaluations are based on uniform methods and
measurements." The Committee's major contribution to
this goal is to attempt to develop a terminology for the
various kinds of evaluation activities, as discussed
above, and to match evaluation questions with appropriate
research approaches. However, we believe that to proceed
any further with specific recommendations for attaining
uniform procedures and measurement is a premature step at
this stage in the development of evaluation.

At the present time, the science and art of evaluation
is in a state of considerable change and improvement.
Each of the social science disciplines has made
contributions to the procedures now used, and while there
is some agreement on the rough preference ordering of
procedures to address a set of policy questions, the
rapid rate of development along with considerable
diffusion of methods from one field to another means that
today's preferences may be superseded by tomorrow's more
mature understanding of the proper fit between problem
and method. In addition, evaluation activities are being
undertaken in a variety of substantive areas--not only in
education, but in manpower training, energy conservation,
health services delivery, child care, public welfare
payment plans, criminal justice procedures, and so

on--and in each of these areas new methods and procedures
are being developed that can be expected to enrich the
field of evaluation.

The Committee believes that, while the goal of
attaining uniformity in evaluation methods and measures
is an extremely desirable one, it cannot be attained at
the present time without prematurely inhibiting further
advances in the field of evaluation and stopping it short
of needed development. The recommendation below that the
National Institute of Education (NIE) continue and
strengthen its program of support for basic research in
evaluation methods is made in part to accelerate full
development of the field of evaluation.

Recommendation D-3. The National Institute of Education
should continue and strengthen its program of support for
research in evaluation methods and processes.

The field of evaluation is a relatively new one that
has made considerable progress in the last 15 years;
however, it is far from fully developed. It continues to
apply promising research approaches from all the social
science disciplines and feed back to them the resulting
experience. Hence, support of research in evaluation
methodology not only improves the field of evaluation,
but enriches the basic disciplines--an effect that is
also important for fundamental research in education.

The Committee believes, however, that support for
development in evaluation has been uneven, in particular,
that too much attention has been given to investigating
problems in the use of randomized controlled experiments,
a procedure that has only limited utility in evaluation
generally. As a result, other important problems in
methodology have not received sufficient study.
Especially important is the development of methods for
studying the delivery of services (implementation), for
investigating the properties of achievement tests when
used in the evaluation of programs (rather than in
ranking individuals), and for assessing the impact of
programs that cannot be studied through the usual
experimental paradigms.

Another neglected area of research has to do with the
process of evaluation itself: how studies are
commissioned and initiated, how they are managed, what
procedures govern their execution, what legal constraints
impinge upon them. Evaluation is controlled by at least

three different agencies: the sponsor of the evaluation, the program or service agency in the field (e.g., a school system), and the evaluators. When the sponsor is a federal agency, there are three control points within the agency: the evaluation monitor, the contracts office, and the manager of the program being evaluated. The complexities created by these multiple organizational relationships create constraints for any study, and those constraints have been given little attention. Our own limited findings related to such issues are reported in the next three chapters; those findings make it clear that the evaluation process must be better understood if it is to yield good results.

The National Institute of Education should encourage work in the noted areas of methodology and process as part of its evaluation research program. Furthermore, with rare exceptions, when a specific methodological question must be addressed in a given time frame or the process of a specific evaluation is to be studied, all such research should be carried out through a competitive grants program that specify the areas of interest but not the approach to be taken.

NOTES

1 Success here is defined in terms of the objectives of the program. It is quite possible that a program successful with respect to its own objectives may be educationally undesirable. For example, perhaps more time was spent on a targeted skill and so some other important skill was neglected and hence less developed than it would have been in the absence of the program. To gauge the overall educational contribution of a program, it is necessary to assess such negative as well as the positive effects.

2 A good deal of knowledge that can be applied to program improvement may, in fact, be gained through documenting program variations and their effects. A panel of the National Research Council's Committee on Child Development Research and Public Policy is currently reviewing outcome measurement in early childhood demonstration programs. Given that local program variation is encouraged by many early

childhood programs, the panel has given considerable
attention to the need to consider the relationships
between variations in treatment and outcomes within
programs and on adaptations in program practice and
variations in outcomes from site to site.

3
Quality of Evaluation

Knowledge about the quality of evaluation studies in
education is limited. It comes from three sources:
technical critiques and reanalyses of specific (usually
large-scale) studies, a few scattered reviews of some
samples of evaluations, and analyses of the inf ence of
the political context on the quality of evaluat ons. The
effects of the managerial context on quality-- ow
evaluations are commissioned and carried out--has
received considerably less attention. Yet the level of
funding, what types of organizations usually perform
evaluation studies, and the availability of adequately
trained individuals all influence the quality of
evaluations. In addition, procurement procedures can
encourage or discourage creativity, and
interorganizational complexities can introduce delays
that often have deleterious effects on the course of a
study.

There are several dimensions to the issue of quality.
Evaluations can be competently done but not be very
creative. They can be imaginatively done but be sloppy
on some points. The various standards for evaluation
work recently developed by a number of groups (Joint
Committee on Standards for Educational Evaluation 1980,
U.S. General Accounting Office 1978, 1979, 1980b,
Evaluation Research Society 1980) may be useful to the
profession, but since any major evaluation is a
customized task, they cannot resolve quality issues in
any specific instance. Furthermore, quality is
inevitably subjective, especially in an activity such as
evaluation for which facts and values are inextricably
linked. For these reasons, the Committee's

recommendations do not feature rigid requirements.
Instead, the Committee has chosen to highlight some
defects that commonly stand in the way of improving the
competence, creativity, and integrity of evaluation and
to propose ways of institutionalizing some quality
control mechanisms. In this chapter, we first review the
available evidence on the quality of evaluations and on
the influence of the political context and then analyze
some of the managerial constraints that affect quality.
In the last section, we focus on evaluation at the state
and local levels.

REVIEW OF THE EVIDENCE

Critiques of Individual Studies

Individual studies of evaluations have generally centered
on evaluations of highly visible programs with strong
advocates and adversaries. Some prominent examples in
education include: the reviews of Equality of
Educational Opportunity (Coleman et al. 1966), which were
edited by Mosteller and Moynihan (1972); the critiques of
the Westinghouse-Ohio study of Head Start (Cicirelli and
Granger 1969), which were initiated by Campbell and
Erlebacher (1970) and grew so voluminous that the
critiques themselves have been analyzed and their impact
assessed (Valentine and Zigler 1979, Datta 1975, 1976);
the evaluations and reevaluations of "Sesame Street" (for
example, Ball and Bogatz 1970, Bogatz and Ball 1971, Cook
et al. 1975); the evaluation of the effects of the
Emergency School Aid Act (ESAA) programs (Crain and York
1976, National Opinion Research Center 1973), which was
then the subject of critiques by the National Advisory
Council on Equality of Educational Opportunity (1975) and
Acland (1975); and the recent evaluation of bilingual
education (Danoff 1978), which has received much
political as well as some technical criticism from the
National Institute of Education and others (U.S. Congress
1977). Both the technical and the political criticisms
have helped the evaluation field to mature, although the
debates have at times been acrimonious and appeared to
confuse rather than illuminate program achievements and
conditions. The debates may also have created a degree
of cynicism about evaluation. Whatever confusion and
disenchantment the critiques and debates have engendered,
however, they have served to sensitize evaluators to

methodological pitfalls and to the need to consider the context in which evaluation takes place. More specifically, as we noted above, they have given rise to several sets of evaluation standards. Unfortunately, the total number of studies subjected to open professional review has been small, and the absence of such review has not necessarily inhibited the use of evaluation findings. Datta (1979) analyzes an interesting example of a study on the effects of federal education programs (Berman and McLaughlin 1975-78) whose summary findings were widely accepted and applied in policy formulation without questioning when later examination revealed considerable problems with some of the summary conclusions and the interpretations they had been given.

Reviews of the Field

Aside from the critiques of some landmark studies, there have been few systematic reviews of the quality of evaluations, such as assessments of representative samples of studies published during a specified time period or resulting from the activities of a particular sponsor or group of performers. In an early study, Bernstein and Freeman (1975) started with 236 studies from fiscal 1970, of which they ruled out 84 as not being comprehensive, i.e., not measuring both process and impact. Using criteria oriented toward quantitative and experimental methodology, they found only 27 of the remaining 152 studies to be of high quality, less than 20 percent; 76, or 50 percent, were deemed to be of low quality. Minnesota Research Systems, Inc. (1976) examined 110 research studies (about 45 percent of which were classified as evaluations) funded by the U.S. Department of Health, Education, and Welfare (HEW) and completed in 1973 and 1974. Less than 10 percent were deemed to be free of significant methodological flaws. Moreover, they found that in 90 percent of the cases the flaws already existed at the proposal stage.[1]

The size and the scale of evaluation studies have grown considerably since the early 1970s, but problems of quality appear to persist. Rossi (1979b) reports on an examination, done over 3 years for the Summer Evaluation Research Institute at the University of Massachusetts, in which several hundred requests for proposals (RFPs) were screened to look for those likely to lead to a sound research plan. Using that criterion, less than a dozen

were identified as being suitable for teaching purposes.
On the performer side, evaluation researchers who
screened more than 100 evaluation research reports on
behalf of the Russell Sage Foundation identified only
some half dozen that merited special review as examples
of high quality (Rossi 1979b). Abt (1979), who heads one
of the major firms engaged in evaluation research, has
estimated that only 5 to 20 percent of studies in the
field of evaluation can be considered valid and relevant
research. He notes that these numbers might be
acceptable compared with those for basic research but
that they are far lower than is the case for other
applied fields such as enginec ng or legal research.
    Indirect evidence on the quality of evaluation studies
comes from a number of attempts, briefly noted in Boruch,
Cordray, and Pion (Ch. 5 in Boruch and Cordray 1980), to
identify exemplary programs. Such attempts--for example,
finding effective programs to increase equity in
vocational education, programs in bilingual education,
and programs in career education--usually yielded only a
small number for which sufficient evidence was available
to make judgments as to their educational promise. The
number of projects so identified tended to be less than
10 percent. Only in the case of the Joint OE/NIE
Dissemination Review Panel, which judges exemplary
projects proposed for dissemination, is the rate of
projects that show adequate data on effectiveness more
than 50 percent; as Boruch, Cordray, and Pion note (Ch.
5:7 in Boruch and Cordray 1980), however, this estimate
is "biased in the direction of higher quality due to
voluntary submissions" and the efforts by the panel to
promulgate its standards for acceptable evidence, which
were published by the U.S. Department of Health,
Education, and Welfare (Tallmadge 1977).
    Except for Bernstein and Freeman (1975) and Minnesota
Research Systems, Inc. (1976), these sources of
information on the quality of evaluation studies do not
distinguish between studies commissioned at the federal
level and those commissioned or carried out at the state
or local levels. A number of the studies commissoned and
funded by the Office of Education's central evaluation
unit have been widely recognized for their technical
proficiency in terms of general standards prevailing in
the field. The picture at the state and local levels is
decidedly more mixed, as documented in two studies cited
by Boruch, Cordray, and Pion (Ch. 5 in Boruch and Cordray
1980) that considered the quality of evaluations

performed at those levels. The first study, by the U.S.
General Accounting Office (GAO) (1977), surveyed state
and local officials on how sound and reasonable they
considered evaluation findings from reports produced by
state and local agencies. While reports issuing from the
same level of government were more credible to officials
(i.e., state officials rated state reports more highly,
local officials rated local reports more highly), even
the most favorable ratings considered only two-thirds of
the reports adequate or better, and in the least
favorable cases (state views of local Title I reports),
barely one-third were considered to be adequate or
better. Among other recommendations, the GAO requested
that the Office of Education review the program
information collected in local agency evaluation reports
in order to determine whether such information could be
aggregated to serve the different needs of federal,
state, and local governments.

In the second study, focused on evaluation carried out
at the local level, Lyon et al. (1978) reviewed 116
studies for the presence or absence of criteria
considered to be necessary elements of an evaluation. As
Boruch, Cordray, and Pion note (Ch. 5:7 in Boruch and
Cordray 1980), the Lyon study "suggests that simple
standards are not often adhered to." Holley (Appendix C)
comments that among the possible reasons are insufficient
evaluation funds, insufficient control of the funds and
often of the evaluation activities themselyes by program
administrators, and lack of training and experience of
many of the personnel who are assigned evaluation
responsibilities.

The Political Context

One of the sources of disappointment with evaluation is
that it appears not to have contributed as effectively as
hoped to the making of decisions about programs. At
times, this lack has been attributed to the inadequate
quality of many evaluations. More recently, however, the
analytic literature dealing with the contributions and
failures of evaluation has reflected a considerable shift
regarding the potential for decision making offered by
program evaluation. Such early studies as the
Westinghouse-Ohio evaluation of Head Start (Cicirelli and
Granger 1969) were in part condemned for a narrow choice
of outcome measures that did not adequately reflect

program goals. More recent writing has emphasized the
diffuseness, multiplicity, and ambiguity of goals in most
social program legislation. Without specification of
outcomes that can be measured, program evaluation as
originally envisaged loses credibility because the
effects achieved cannot be compared with those intended.

Researchers do not agree on how to deal with the
dilemma of program legislation that may be specific on
process but is vague on intended objectives, yet mandates
evaluation. Rossi et al. (1979) have suggested that
program goals should be spelled out specifically enough
to allow impact assessments; more recently, he and Chen
(1980) have argued that researchers cannot simply accept
official goals but must learn how to interpret programs
and their likely effects more accurately in order to
design evaluations that are sensitive to program impact.
Wholey, when he was Deputy Assistant Secretary for
Evaluation of HEW, introduced the notion of evaluability
(see Appendix A) whereby short-term, exploratory
evaluations would determine the operational objectives of
a program and whether they could be measured (Wholey
1979); if they could not, costly impact assessment would
not be commissioned. Cronbach et al. (1980) argue that
the quest for specification of goals is futile and that
evaluation is a prospective activity better suited to
understanding processes and events for future program
formulation than for retrospectively appraising the
performance of programs against predetermined objectives.

There is more agreement on the role of the evaluator
in the decision-making process, namely, that the
information developed through the processes and by the
canons of social science is, and should be, only one of
the determinants of policy regarding education (or any
other social) program decisions. Arguments deriving from
research on how evaluation findings are used (Caplan
1977, Alkin et al. 1979) have led to recommendations that
evaluations, to be useful, must be done in close
cooperation with the intended user and must also involve
a process of negotiation that draws on the views of
beneficiary and constituency groups. However, such a
process is often counter to the objectivity considered to
be a hallmark of quality evaluation. According to
Schreier (1979), it pits the insider's (e.g., client's,
teacher's, program manager's) intuitive perception
against the outsider's concern with quantitative
assessment. The result is that they are unlikely to
agree on goals. The focus of evaluation may then shift

to good management, the purpose being to improve program
process rather than to ascertain how well outcomes, which
remain unspecified, are being met.

THE MANAGEMENT CONTEXT

Over the last decade, evaluation of education programs
has become big business, and this has had an impact on
quality. When the first legislative mandate for
evaluation was written into law as part of the 1965 Title
I (ESEA) legislation, evaluation was considered to be an
activity carried out at the local level for
accountability and to improve the program. Every year
thereafter, local evaluation activities were initiated
for a number of programs, usually coordinated by an
evaluation specialist within the federal program office.
As the number of activities grew, concern with quality
and need for generally applicable procedures led to the
establishment in fiscal 1970 of a central evaluation unit
in OE (see Appendix A).

Funding

Before fiscal 1970, the Office of Education had about
$1.25 million per year for central evaluation available.
In that year, for the first time, there was a separate
line item for evaluation. The peak funding for the
central evaluation unit was reached in 1978, with $29.7
million obligated for evaluation contracts. In 1980, the
amount had decreased to $19.4 million. The most
precipitous drop within the unit came in evaluation funds
for discretionary purposes, i.e., not earmarked for a
specific title: these funds dropped from $7.1 million in
1977 and 1978 to $3 million in 1980 (U.S. Department of
Health, Education, and Welfare 1979b).
   According to Reisner's estimate (Appendix A), in
fiscal 1980 the Department of Education was planning to
spend some $40 million on a variety of evaluation
activities, half of the work being carried out by the
central evaluation unit and nearly a quarter by the
Inspector General. If one wishes to calculate the total
amount spent for program evaluation in education, that
estimate needs to be augmented by the amount spent by the
General Accounting Office (estimated at $2.5 million) and
an unknown amount of federal funds devoted to evaluation

activities carried out or commissioned at the state or
local levels. Taking a different approach, Sharp's
analysis (see Appendix B) is based on performer rather
than sponsor data, includes policy studies as well as
evaluation activities, and is for 1977 when there may
have been a somewhat greater investment in evaluation:
her best estimate is that a total of $100 million in
federal funds was spent for evaluation in education at
all levels of government. This amount represents
something like a fourth or a fifth of all evaluation
activities funded by the federal government. By far the
largest growth occurred during the earlier part of the
decade (see Abramson 1978, Cronbach et al. 1980, National
Science Foundation 1979); during the last few years,
federal funding for evaluation, at least that portion
visible at the national level, has actually decreased
somewhat, matching the trend for overall funding for
education. As a percentage of total federal expenditures
for education, the current investment in evaluation
represents about 0.5 percent of the total federal support
for education, which stood at $14.2 billion in fiscal
1980.

Performers

Although expenditures for evaluation may appear modest as
a percentage of expenditures for education, they are a
major source of income for private-sector performers of
educational research and development.[2] Such performers
account for nearly half of the total spent for evaluation
(Appendix B:Table B-4) and are particularly prominent in
carrying out medium-scale ($100,000-$500,000) and
large-scale (more than $500,000) studies
(Appendix B:Table B-5). Within the private sector,
for-profit firms report that more than 50 percent of
their research activities consist of evaluation and
policy studies (Appendix B:Table B-3). By contrast, less
than 10 percent of academic institutions carry out
medium- or large-scale studies; some 40 percent report
doing no evaluation work at all (Appendix B:Table B-5).

Moreover, evaluation work is heavily concentrated
among major[3] private-sector performers; they account
for 83 percent of evaluation funds spent in the private
sector (Appendix B:Table B-8). They are also more
heavily dependent on federal funding than any other set
of institutions (Appendix B, Table B-9). As Sharp notes
(Appendix B:219):

Large private-sector organizations and
organizations which specialize in education RDD&E
have especially few other sources of funding: half
of the organizations which expended more than $1
million in 1975 for education RDD&E received at
least 75 percent of their funds from the federal
government, and one fourth of them received at
least 90 percent from this source.

Personnel

Evaluation is a relatively new field that is to a
significant degree staffed with individuals recruited
from other fields. This newness creates a critical
quality problem at the state and local levels (see
below), but important gaps exist throughout the
evaluation enterprise. Of specific concern are the
underrepresentation of minority group members in
educational evaluation, the communication barriers
between evaluators and administrators, and the failure of
individuals charged with evaluation responsibilities to
keep up with developments in the field.

Toward Equal Educational Opportunity

In order to further the national commitment to equal
educational opportunity, nearly 80 percent of federal
education programs are targeted for racial, ethnic,
handicapped, and other minority or disadvantaged groups.
And if federal programs are to provide more effective
educational services for these groups, consistent input
on their needs must be part of the evaluation process.
An examination of social science research over the last
40 years (Gregg et al. 1979) shows how research questions
have changed in those fields--and those fields only--in
which the subjects of inquiry have participated actively
in defining the problems. Though talent and skill remain
the prime requisites for evaluation personnel, the
perspective that comes from being a member of the
recipient group augments the evaluation process in
important ways. Thus, one can look at bilingual
education from the viewpoint of society as a whole, of
the classroom teacher, or of the non-English-speaking
child and family. Women, blacks, and other minorities
have helped give a different cast to educational research

that arises out of their perspectives. For this reason,
the Committee is concerned that individuals from these
groups who could contribute to broadening evaluation
perspectives are not adequately represented in the
current staffing and procurement of evaluations.

For example, of the 65 professional staff of the
central evaluation unit of the Department of Education in
March 1980, there were 4 black men, 2 black women, 1
Asian man, and 19 white women. There were no Hisp....'.s
or American Indians on the staff. For another example,
in the technical assistance centers (TACs), which have
been created to aid local projects in conforming to the
guidelines and standards set for Title I evaluations and
which presumably should act as models for expanding the
audience and decentralizing the process, not a single
director or senior staff person was a minority individual
as of spring 1980. Of more than 100 evaluation
professionals at any level in the TACs, there were only 8
minority persons. Principals in the central evaluation
unit have consistently expressed a desire to hire more
ethnic and racial minority persons in key professional
positions, but, according to them, have not been
successful in finding those with the appropriate
background and necessary skills.

As a group, minority-run firms have fared particularly
badly in the field of evaluation. Despite special
provisions for 8-A contracting,[4] only 15 of 200 new
contracts awarded by the central unit during fiscal 1976
through fiscal 1980 went to minority firms,[5] 8 through
the 8-A process and 7 through the competitive process.
These 15 evaluation contracts accounted for less than $4
million of a total of close to $100 million awarded in
those years, or barely 4 percent of the total, and only
10 minority firms were involved.

The issue is not simply nor even primarily an
affirmative action one. We presume that both the
Department of Education and its contractors and grantees
are complying with the laws regarding equal employment
and affirmative action programs. In fact, it has been
argued that women and minorities are already represented
on staffs and in the evaluation enterprise proportionate
to their percentage in the available talent pool. But
this is not the only criterion: they are greatly
underrepresented compared with their numbers in the
beneficiary population. The Committee is not suggesting
proportionate representation, but we are stressing the
importance of this issue in personnel and procurement

practices. In our recommendations below, we suggest some
means for greater involvement of minority firms and
individuals in performing and reviewing evaluations. Our
first recommendation addresses the issue of the talent
pool, since unless it is expanded minority participation
in evaluation will continue to remain limited. At the
same time, the recommendation considers some additional
gaps in the training of evaluation personnel that must be
remedied if the quality of evaluations is to improve.

· Training

Recommendation D-4. The Department of Education should
provide funds for training programs in evaluation to
increase the skills of individuals currently charged with
carrying out or using evaluations and to increase the
participation of minorities.

This recommendation covers three training needs that
require extramural support: recruitment and training of
minority individuals; training to improve the
communication between evaluator and the user of
evaluations; and training for those currently involved in
evaluations. Two related issues are covered in other
recommendations: broader technical assistance to state
and local agencies is discussed later in this chapter,
and intramural training for federal evaluation and
program staffs is discussed in Chapter 5.

After 15 years, the rationale that there are no
minority researchers available to help evaluate education
programs is not tenable. Their absence is particularly
marked, and particularly detrimental, at the senior
levels of both sponsoring and performing organizations.
There are increasing numbers of minority persons in
training in Ph.D. programs in social and behavioral
sciences, in part because of numerous federally sponsored
fellowship programs.[6] These social and behavioral
science graduate students very often express interest in
"applied research," but do not often have an opportunity
to learn about it. They represent a sizable pool of
potential evaluation researchers who could staff
positions in the Department of Education, who could
advise and consult with local and state evaluation
groups, and who could work with universities and private
consulting (including 8-A) firms in carrying out
evaluations. Fellowship and internship programs in

evaluation that include specific priorities for minority
group persons would be doubly valuable. They would
produce good researchers and they would enrich the
evaluation system. Some of the current fellowship
programs could include a special component for people
studying evaluation, and internships could be made
available for people in their third or fourth year of
doctoral study. Such internships might be coordinated
through contractors, states, or local school systems
doing evaluations of federal education programs. A
percentage set-aside from evaluation contracts might be
used as a pool of money for mounting such a national
program. Alternatively, RFPs or grant announcements
might require that such internships be budgeted and the
training parameters specified. A feeder system through
other federal fellowship programs concerned with
increasing minority participation in social science
research and development activities could also be
initiated.

     The second training need concerns the relationship
between the evaluator and the administrator or educator.
There is often a communications gap between the two that
renders the use of evaluation far less effective than it
could be. This gap might be narrowed by appropriate
training on both sides. Executives and program staff
could benefit from greater knowledge of the language of
evaluation and how evaluations can be used. Short
training sequences on such topics might be developed and
made routinely available to new staff. For the
evaluator, who often lacks experience in program
management or delivery, exposure to the problems,
procedures, and constraints of federal education programs
would be similarly beneficial. In addition, training
should be directed to improving both the interpersonal
and the communication and reporting skills of the
evaluator so that evaluation information is conveyed as
usefully as possible.

     A third type of training is needed to assure a
minimally adequate level of skills for persons newly
assigned to evaluation responsibilities and to allow
others to keep up with the field. Despite the entry into
the field of many individuals without the requisite
skills and the rapid development of evaluation
techniques, which makes once-adequate skills obsolete,
training in evaluation training is currently inadequate
or unavailable. The Committee is less interested in the
number of new graduate students recruited to the field

than in improving the skills of current performers and
users. Sufficient numbers of staff trained in either
rigorous evaluation methods or in research have never
been available. As a consequence, evaluation is
currently practiced by people with almost every type of
background possible, including many with no more
preparation than that of classroom teaching. These
practicing evaluators need opportunities to upgrade and
improve their skills. (See Appendix C for details on
training needs among local personnel and on some possible
programs.) Insofar as new evaluators continue to be
recruited, graduate-level training programs for
evaluators will continue to need support. In part, such
training would occur automatically through greater
participation of the academic sector in evaluation work
sponsored by the Department.

The suggestions in this recommendation require the
funding of extramural training and fellowship programs.
One channel for such programs might be the Assistant
Secretary for Educational Research and Improvement,
either through the Office of Dissemination and
Professional Improvement or through the National
Institute of Education, which already runs a program to
increase the participation of women and minorities in
educational research and development (R&D).
Congressional authorization for such programs already
exists, at least for NIE, in the 1980 Higher Education
Amendments (P.L. 96-574), and in the Special Projects
Act, though the latter requires that Congress be notified
before a program is initiated.

Interorganizational Complexities

There is an important difference between most social
science research and evaluation. In most research,
control of a study is mainly in the hands of the
researchers: they decide what to study and how the
research is conducted. Even when action sites like
schools are involved, the researchers select them on the
basis of the intended research design, and if some sites
are unwilling to cooperate, others can be substituted.
The funding agency's role is usually limited to
negotiating grant amounts and requiring nominal progress
reports.

In evaluation, the researchers share control to a
considerable extent with two other parties--the

sponsoring agency and the program or action agency.
First, the sponsor sets conditions by designing the RFP
that solicits the evaluation, including the level of
effort, the scope of work, the types of issues, the
research design and measures that are to be used, and the
timing. Second, the nature of the action program itself
imposes constraints, including how funds are allocated
within the program, how far along it is in the
implementation process, how much freedom is given to
individual sites to carry out their own miniprograms.
Third, the research team must work with a specific set of
action sites. In order to establish workable
relationships with action sites that may be reluctant
participants, the researcher must provide a set of quid
pro quos, such as collecting data not necessarily
relevant to the evaluation study but wanted by people at
the site, providing technical assistance, or carrying out
special analyses. Moreover, neither the action site nor
the sponsor is a monolithic entity, and different
requirements and constraints may be imposed by different
organizational units within each. Of particular
importance is the increasing fragmentation of
responsibilities within federal executive agencies (the
usual type of sponsor), in which at least three parties
may have some influence over the design and conduct of
research: the project monitor for the evaluation study
itself (and the cognizant evaluation unit), the program
manager and responsible office for the program being
evaluated, and the contracts office. The resulting
context for evaluation is depicted schematically in
Figure 2 (see Yin 1980).

Ine quality of evaluation is subject to the marked
constraint imposed by the need for researchers to work
within these interorganizational complexities: each
decision has to be negotiated and agreed to by a number
of parties. If nothing else, the process of arriving at
compromises acceptable to all parties is time-consuming,
often to a degree that makes the original study design no
longer feasible; this is especially true during the
procurement phase and the implementation phase.

The low participation of the academic sector in
evaluation work should not be surprising, even though
academic organizations represent the largest single group
of performers of all educational research (Appendix
B: Table B-4), because of the process by which evaluations
are procured by the federal government. That process has
become more and more complex over the decade of growth in

```
                    Sponsoring Agency
                  (usually a federal agency)



    Evalu... on Project         Contracts              Action Program
    Monitor                     Office                 Monitor



    Research Team                                 Action Agency(s)
    (usually a university or                      (usually state or local
    independent research group)                   units of government)
```
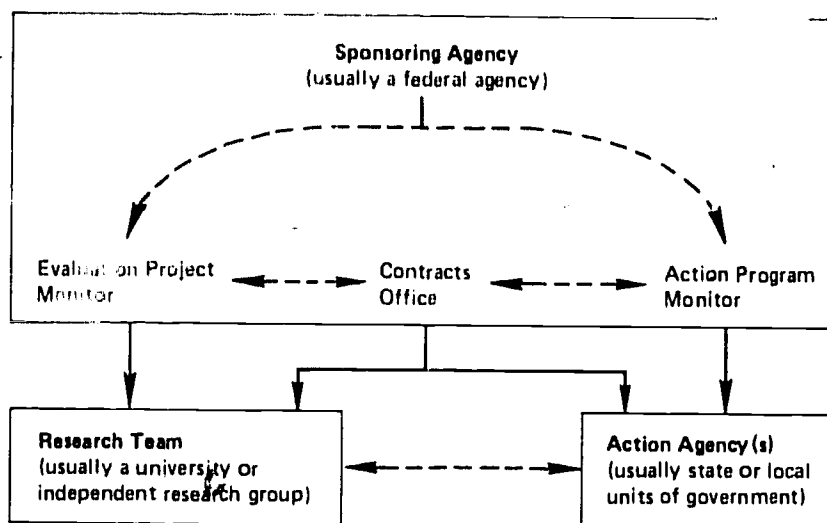
FIGURE 2 The interorganizational complexities of
evaluation research.

evaluation funding. Requests for proposals (RFPs) have
become longer and more detailed: in addition to spelling
out basic design, methodology, what to measure and how to
measure it, they may specify the sites to be studied, the
data elements to be analyzed, and the time intervals for
different collection steps. Responders have little
freedom to formulate research approaches they consider
more appropriate, let alone to reframe evaluation
questions. Moreover, the average response time allowed
hardly permits such luxury: for eight of the ten RFPs
issued for new studies in fiscal 1980 by the Office of
   gram Evaluation (the central evaluation unit for the
 partment of Education), proposals were due only 1 month
after issuance of the RFP; for the other two RFPs,
proposals were due in 6 weeks (see Table 1). The
proposed length of time for these studies ranges from
18 months to 2-1/2 years and their projected cost ranges
from $150,000 to $2 million. The largest of these
studies, which comprises a whole series of substudies of
the implementation of Title I at the state and local
levels, is estimated to take 2-1/2 years and cost $2
million. The RFP for this study was issued on July 23;
proposals were due 29 days later, on August 22. [7]

TABLE 1　Milestone Dates--Fiscal 1980 RFPs--Office of Program Evaluation

	Work Statement First Draft to GPMD[a]	Work Statement Final Draft to GPMD[a]	RFP Issued	Proposals Due (Closing Date)	Contract Award
Development of bilingual evaluation models	1/31	2/13	3/6	4/7	6/27
Assessment of Women's Equity Act Program	2/12	2/22	3/14	4/14	6/30
Description of state management practices in ESEA Title I	3/12	3/24	4/25	5/27	6/30
Assessment of the Strengthening Developing Institutions Program	3/31	6/3	6/25	8/5	9/30
Evaluation of Basic Skills Improvement Program	3/26	6/5	7/1	8/4	9/30
Management studies of federal education programs	4/9	6/20	7/10	8/22	9/30
Evaluation of impact of Part A of Indian Education Act	4/30	6/25	7/18	8/18	9/30
ESAA-funded activities and Management Information System	5/27	6/13	7/3	8/4	9/30
Description of ESEA Title I district programs since 1978	5/30	7/1	7/23	8/22	9/30
Assessment of ESEA Title I Program for Handicapped	5/30	7/7	7/23	8/26	1981[b]

[a]Grant and Procurement Management Division.
[b]Originally planned for 9/30/80, postponed until fiscal 1981.

Tight timetables for preparation of major evaluation proposals are the rule, though the reasons vary from year to year. In 1980, the cause was a complicated internal planning process combined with the need to expend evaluation dollars during the fiscal year in whioh they were appropriated. Evaluation plans submitted by the Office of Program Evaluation in the spring of 1979 were not approved until January of 1980; some studies were not approved until May. Therefore, except for two RFPs that had been held over from fiscal 1979, no work statement could be completed until March, and a number were delayed until June or July by further review within the Grant and Procurement Management Division, the Department's contracts office. Thus, seven of ten planned awards for new studies were not scheduled until September, at the very close of the fiscal year.

Institutions whose business is based on federal contracts resulting from RFPs and who have considerable staff resources assembled at any point have an obvious advantage when responses must be made in such a time frame. The recent change in the federal government's fiscal year has positioned many complex procurement actions in the summer quarter, a period during which academic institutions are even less likely to be able to respond quickly. Contract records substantiate Sharp's findings (Appendix B) that universities and small-scale performers are largely shut out of the types of studies ($100,000 and over) that have been in favor. Of 84 contracts for evaluation and planning awarded by the central unit in 1979, only 1 went to a university, in the amount of $350,000 of a total of $21,526,089 in awards. On the other hand, one for-profit firm received four contracts for a total of more than $5 million. Nineteen contracts to three private firms and one large regional laboratory (also a private corporation)[8] accounted for 50 percent of all funds awarded. Through their success in responding to evaluation RFPs, the private performer organizations have been able to accumulate "large, sophisticated, multidisciplinary staff which are very knowledgeable about the major educational issues of the day" (Sharp, Appendix B:241). Whether current procurement procedures with their tight deadlines and enormous response burdens serve to deploy effectively the talent pool in even this limited domain is open to question. The reviews of evaluation proposals cited earlier in this chapter are not reassuring about the quality of responses elicited by the procurement process.

Constraints operate not only during the procurement
process and original design phase, but also during the
execution of any study. The first obstacle after a study
is launched is to obtain clearances for data collection
instruments. Clearance procedures (described in greater
detail in Chapter 5) may take 5 to 6 months. Three of
four different bodies are involved in the process,
looking at the study design, the data collection
instruments, and the analysis plan from a variety of
perspectives: burden on respondents, technical quality,
need to know (defined as being required by law), and
economic impact. Not infrequently, research designs and
instruments that are the product of experts and that have
been pilot tested are changed by reviewers who do not
have equivalent expertise or field experience. If a
study is to be done at all, many compromises have to be
made along the way by the contractor and federal monitor.

In 1978, a new requirement was added to the clearance
process, namely, that all test and data collection
instruments to be used in a study must be described in
the Federal Register (and available on demand) by
February 15 previous to the school year in which the
information is to be collected.[9] This requirement,
when added to all the other clearance machinery, so
compresses the time available for development of
instruments and questionnaires that quality takes a back
seat to doing the study at all. It also severely limits
the possibility of making changes as a result of
conditions in the field or as promising lines of inquiry
develop during the course of a study. The added costs
engendered by keeping key staff who are essentially
unproductive as they await clearance to go into the field
squanders time and money that could have gone into
improved design, data collection instruments, and
analysis.

Even past the hurdles of clearance, a funding unit
exercises great influence over the nature of evaluation
studies through the monitoring process. When unexpected
conditions arise that may require changes, such changes
will be affected by agency officials because of their
active role in approving or rejecting requested
modifications. Decisions may be slow in coming, since
most of them will require agreement among the three
internal agency parties involved (evaluation monitor,
program manager, and contracting officer). Agency
officials and performers have to understand and resolve
the tension between necessary changes in direction and

timely delivery of an evaluation study; creative skills
are required to negotiate such tensions successfully
without impairing the quality of the study. In some
cases, it may be more important to deliver findings on
time than to ensure that the results are as
methodologically rigorous as possible. The balance
between adequate agency procurement and monitoring
procedures and creativity needed from the field to
produce high-quality evaluations has in recent years
swung heavily toward agency control and, within the
agency, to control by contracts and grants management
specialists rather than by .technical evaluation staffs.
The three recommendations below are aimed at introducing
greater creativity and competence into the evaluation
process during three stages: procurement, while a study
proceeds, and after completion.

Recommendation D-5. The Department of Education should
structure the procurement and funding procedures for
evaluations so as to permit more creative evaluation work
by opening up the process and allowing a period for
exploratory research.

The increasing constraints imposed as a result of the
greater visibility of evaluations and the attempts to
control their management and process have limited
contributions from the field of evaluation. These
constraints have reduced the opportunity for infusing
novel approaches into either programs or evaluations.
They have also reduced the potential of evaluation to
contribute to the policy process.
The more complex the evaluation, the less likely is it
that anyone can spell out ahead of time the best methods
for addressing the questions that the evaluation is
designed to answer. The current RFP process in
particular ignores this fact. The Committee believes
that this process can be made more flexible. An RFP
often presumes some things about the program are known
when they are not. This can range from something
fundamental--e.g., existence of the program at a site--to
something trivial--e.g., existence of records. RFPs also
often downplay the possible effects of interorganiza-
tional relationships on the evaluation process. In
addition, problems and issues in executing the evaluation
are not anticipated, and many cannot be anticipated. The
unknowns or unknowables suggest that an RFP that attempts

to be specific is bound to be inappropriate. Therefore
RFPs should include a period of exploratory research
before the evaluation is undertaken in order to frame
questions properly--with the aid of the consultation
process suggested in Recommendation D-6 below--and to
figure out what the unknowns are. RFPs should also
provide for side studies that are research oriented to
illuminate questions that emerge during the evaluation
and that should be answered if the evaluation is to be
done well.

Precedents for encouraging exploratory research befor
an evaluation is undertaken exist: James Coleman had the
benefit of 1 year of planning for his national
longitudinal study of the high school class of 1980
(Coleman et al. 1979). That planning included intensive
research on what kinds of policy issues could be
addressed in the future using such data. As another
example, the NIE compensatory education study (National
Institute of Education 1976) had 6 months to clarify
questions before the study was initiated.

Mechanisms for providing opportunity for expertise in
evaluation to improve the quality of evaluations include:

   • inviting bidders to specify alternative methods
of evaluating the program at hand and how such methods
would be tested, in addition to asking that they meet
formal RFP requirements;
   • inviting bidders to design small side studies
that can lead to durable general statements about
particular approaches and providing support for those
side studies found to be meritorious;
   • assuring that sufficient time is available for
developing proposals for an evaluation project, at least
6 months for complex evaluations;
   • issuing RFPs for pre-evaluation assessments that
define the problem better, lay out alternative approaches
to evaluation and how they might be assayed, and so forth.

Beyond improving the RFP process, there are other
steps the Departme: t should take to introduce greater
creativity. The procurement process now used by the
Department to obtain most evaluation studies virtually
limits all contract applications to organizations that
have the capacity to assign full-time specialists who can
be immediately responsive to RFPs. Under this system,
the evaluation program is effectively cut off from the
academic community, which has made major contributions to

the theory and methodology of evaluations. It has been
argued that academic researchers are disinterested in
applied research such as evaluation, since they are more
highly rewarded for basic research, and that the
disciplinary structure of universities does not lend
itself to policy-relevant research. Though there is some
justification for these views, one cannot conclude that
universities will not and should not participate in
carrying out evaluations. The academic world is no more
monolithic than any other community; within many
universities, there are institutes or centers created
precisely to respond to the interdisciplinary challenges
of applied social science research. In addition, as
funding for basic research has leveled off or even
decreased, academic researchers have become more
interested in applied work. The dismal statistics on
lack of participation by universities in evaluations
funded by the Department cannot be attributed solely to
the unwillingness of universities to participate.

By depending almost entirely on the competitive RFP
procurement system, the Department is not able to take
advantage of the creativity, objectivity, long-term
commitment, and the cumulative knowledge and experience
of the academic community. Nor can it attract
participation by minority researchers, whose perspectives
would enrich the questions and methods of evaluations,
who are not able to assemble the resources needed for
large studies in the time provided. Local and state
agencies also cannot often contribute at the national
level, even when they have the capability for high-
quality work, because of the site requirements in many
RFPs. Among the mechanisms for funding evaluations that
can be used to open up the process and improve quality
are unsolicited proposals, sole-source awards, 8-A
contracting, cooperative agreements,[10] basic ordering
agreements,[11] and grants.

The Department should consider unsolicited proposals
in order to encourage creative and innovative ideas that
may be lost through the RFP system. Academic experts who
have made significant contributions to the evaluation
process should be encouraged to submit proposals that
attempt to break new paths in theory or measurement of
the effectiveness of education and other social
programs. It is possible to carry out a competitive
program of grant awards for unsolicited proposals in
specified areas, as practiced by the National Institute
of Education.

When the Department wants to take advantage of the
expert knowledge of an academic scholar who may have made
a significant contribution to a particular subject area,
it should have the authority to solicit a specific
proposal. Some members of the academic community have
unique knowledge and skills that are not found
elsewhere. The Department should have the authority to
offer a sole-source award to a scholar in the field of
evaluation whose background, experience, and expertise
cannot be matched. The use of this mechanism will help
to open up the system to new ideas and contribute sorely
needed flexibility to the Department's evaluation
activities. The Committee is fully aware of recent
criticisms of consulting and sole-source procurement
(U.S. General Accounting Office 1980a, Gup and Neumann
1980, but see Wilson 1980). We believe, however, that
the limited and judicious use of this mechanism can
produce gains that far outweigh the risk of occasional
abuse. When abuse does arise, it should be dealt with on
a case-by-case basis, not by abandoning a useful
procurement mechanism.

The restrictiveness of the RFP process also
contributes to the very low use of minority firms by the
Department in securing evaluation contracts. Such firms
are usually small and have limited staff and so they
cannot respond as quickly to RFPs as the larger
for-profit organizations that now dominate the evaluation
field. The 8-A contracting process seems to be seldom
used as a way of involving more minority firms, probably
because evaluation studies have tended to be large scale
and 8-A firms are small. The issue of equal educational
opportunity that calls for the greater use and
involvement of minority researchers will only be resolved
when more flexibility is built into the design of studies
and the contracting process.

Cooperative agreements ought to be the mechanism of
choice when the principal purpose of the award is to
benefit local or state operation of education programs
authorized by federal statute. Such agreements may also
be used when substantial involvement is anticipated by
the federal agency as well as by the recipient of the
funds. Studies carried out by a state or local agency to
document program processes, improve program
implementation, or test program alternatives are intended
to benefit the locality, but they can also help improve
the program nationally. The former Department of Health,
Education, and Welfare had an internal decree against

cooperative agreements, though they are used by such
other agencies as the National Science Foundation and the
Law Enforcement Assistance Administration of the
Department of Justice. The Department of Education
should exploit the potential of this procurement
mechanism. Cooperative agreements are an obvious vehicle
for encouraging local and state agencies that have the
capacity to undertake evaluation work aimed at program
improvement.

Basic ordering agreements are a particularly useful
mechanism for planning or evaluability studies and other
limited work with a short time horizon. The Department
could obtain greater flexibility and faster turn-around
time by maintaining lists of qualified performers
generated through periodic requests for qualifications
(RFQs). These performers could then be called upon for
limited studies.

Grants are a particularly appropriate mechanism when
creativity from the performer is important. The
Committee urges that the Department institute at least
two grant programs, one for local and state agencies (see
Recommendation C-3 below) and a small grants program
($50,000-100,000 per grant) to allow university
researchers and others to pursue evaluation questions in
designated areas of interest to the Department. The
small-grants program should be run in conjunction with
the research program at NIE suggested in
Recommendation D-3 (in Chapter 2). Research grants are
often considered to be appropriate only when the primary
audience is to be other researchers and hence are
considered inappropriate for policy-related reseach. But
grant programs do not have to be untargeted, as is
demonstrated by the well-defined grant programs developed
by the various study sections of the National Institutes
of Health and of Mental Health. Not infrequently, the
research is both applied and immediately applicable, as
in the case of the restorative materials program funded
by the National Institute of Dental Research.

The state and local program we are recommending could
be in the form of grant awards or cooperative
agreements. The purpose would be to allow selected
agencies to study their own federally supported programs
by documenting what actually goes on in the program at
the classroom or school level, assessing the effects of
the program or some of its components, and testing
alternative program interventions. There should be
national or regional competitions for each large federal

title and one catch-all category for the small programs.
Panels of outside experts (including nonfederal
researchers) should evaluate proposals. Proposals should
be required to state how results of a study will be
incorporated into pertinent local or state agency
operation. The Department should use existing mechanisms
like state agency dissemination arms, assistance centers
attached to various federal education programs, or the
National Diffusion Network (NDN) to disseminate and apply
findings nationally.

The Committee's recommendation that a greater variety
of procurement methods be employed does not suggest that
the use of RFPs be drastically reduced. We recognize the
need for organizations that can m unt nationwide surveys,
carry out complex tasks, and have available large numbers
of experienced analysts. Our call for flexibility in the
procurement process, we believe, will reduce the
sterility of the evaluation system through the
introduction of new ideas and will permit increased
consideration of different perspectives that can
contribute to the educational system.

Review

A common defect in past evaluations has been that only a
small group of people in the agencies and among the
contractors are talking to each other; they are doing
things in standard ways and perhaps missing new
developments in technique or new ways of evaluating or
running programs. The results of evaluations are then
made available and often taken on faith by the
educational community. Since evaluation is a difficult
and ambiguous activity, the evaluation process would, in
the Committee's view, be improved by opening it up--even
if this results in longer time frames.

Recommendation D-6. All major national evaluations
should be reviewed by independent groups at the design,
award, and final report stages. Review groups should
include representatives of minorities and other consumers
as well as technical experts. The results of their
review should be made broadly available.

Insofar as it is feasible, such reviews should also be
conducted for major state and locally sponsored
evaluations.

This recommendation has three facets to it: improving the technical quality of evaluations, ensuring early contribution and involvement from those most affected by the programs (beneficiary groups, teachers, etc.), and making use of the findings more likely through public exposure and understanding.

For major national evaluations of important programs, the evaluation plan should be publicized by the agency before the project begins. When the RFP process is used, the agency itself should solicit as much outside advice as possible, thorough development of concept papers, planning conferences, and other pre-RFP activities. Proposal review should include experts from outside the sponsoring agency. After award of a contract, the contractor also should solicit the views of outsiders. Some questionable assumptions or pedestrian analytical approaches might be amended at this point. Then, when the project is done, outsiders should again review the work, its philosophical perspective, its technical ambiguities, and its policy implications. Such outside review would be facilitated if researchers were careful to spell out, in final reports, the limitations of their research: ". . . what went wrong, what couldn't be done, what that means for the conclusiveness of the findings and . . . for their generalizability to particular populations" (Chelimsky 1978). 'Later on, the data from the evaluation should be made available to others for reanalysis. If evaluations are controversial, either because of their execution or because of their recommendations, this process will allow such controversies to be aired. All of the results of this interchange, the evaluators' reports and the comments of outsiders, should then be made broadly available.

There may be several ways to ensure adequate input and broad availability. One approach worth exploring is for the Department to sponsor an annual conference on important evaluations that are at various stages in the process--design, first completion, reanalysis. If this were done, the educational community would know where to look for the latest evaluation results, criticisms, and reanalyses, as well as for information about impending work.

In line with previous remarks about the subjective nature of evaluation quality, opening up the evaluation process should provide mechanisms similar to those employed by such journals as Consumer Reports with regard to the market for consumption goods. The Department

shr ld not be the arbiter of evaluation quality. But it
can make sure that all evaluations are subjected to the
scrutiny of outsiders so that the educational and
beneficiary communities at large, as consumers of
evaluation information, can see the pros and cons, the
ambiguities and questions, and make up their own minds.
In the long run, this greater information and exposure is
the surest way to make certain that evaluations will
consider the perspectives of parties at interest, will be
of high quality, and will not be ignored.

This recommendation implies that evaluations will not
generally result in an immediate consensus on the value
of an education program. To a certain extent, this lack
of consensus is a fact of life in the field of education,
and the Committee would be remiss if it did not warn
Congress and the Department of Education of this fact.
But we see in the suggested mechanism some ways of trying
to resolve the real controversies. As part of a
subsequent reanalysis process, conference participants
might try to agree in advance on further analyses to be
done and what they could show. In that way, there might
be a greater chance of arriving at agreement on the
results of the second round of tests and analyses. The
same logic also applies to the idea of presenting
evaluation plans: it is likely that when more voices are
heard early on, less acrimony will be heard later on.

Recommendation D-7. All statistical data generated by
major evaluations should be made readily available for
independent analysis after identifying information on
individual respondents has been deleted.

When possible, ethnographic data and case study
material, similarly treated to protect privacy and
confidentiality, should also be made available.

The data generated in most large-scale evaluations are
an expensive resource and should be treated as such.
They can be reanalyzed in the interests of critical
appraisal of the original evaluation and in the interest
of advancing the theory of program testing and the state
of the art in evaluation. They can be useful for
pedagogical purposes in university training and for staff
development in government and in state and local
education agencies. Mechanisms for ensuring that the
data are available for reanalysis include: provision of
support for documentation, storage, and dissemination of

data in major evaluation contracts; creation of explicit
agency policy on access to data; and statutory
requirements for independent review and, where
appropriate, reanalysis of original evaluation data.

Independent reanalysis of data generated by
evaluations should capitalize on procedures that avoid
compromising the privacy of individuals or the
confidentiality of information. Audit agencies such as
GAO, or independent researchers, may have a legitimate
interest in verifying quality of data generated in an
evaluation. The process need not and should not breach
promises of confidentiality made to individual
respondents or invade their privacy. A report
commissioned by the GAO on assessing evaluation quality
(Social Science Research Council 1978) recognizes the
additional needs of avoiding needless disruption of
research and harassment of respondents. The report
recommends several alternatives to the usual way of
reinterviewing respondents including: independent
sampling of the target population to compare statistical
results obtained by the auditor with statistical results
obtained by the evaluator; use of evaluators independent
of both original evaluation staff and audit staff for
reinterviews; drawing a subsample of the original sample
for reinterview to minimize disruption of the research;
and other strategies. In many intances, regathering of
primary data is unnecessary: review of design,
execution, and analysis is sufficient for judging the
quality of major program evaluations (see also Hedrick et
al. 1979). The critical point is that original
evaluation information not be withheld by researchers,
sponsors, or any other parties; the more such information
is available, the less intrusive can be the approach
taken in reanalysis and critical appraisal.

STATE AND LOCAL ACTIVITIES

Funding and Independence

The amount of federal money spent for evaluation
activities at state and local levels is not
inconsiderable. Webster and Stufflebeam (1978; see
Appendix C:Figure C-3) found that 35 large urban school
districts spent a total of nearly $34 million on research
and evaluation, of which $21 million (or more than
two-thirds) was federal funds. But funding for

evaluation varies widely. The size of local education
agency (LEA) budgets for the evaluation of Title I
programs has ranged from 0 to nearly $1 million for
programs that have a total budget of more than $100,000
to $52 million, respectively (Drezek et al. 1980; see
Appendix C). There is also great variability for
different programs: for example, an average of 1 percent
of program funds is spent at the local level for
evaluation of P.L. 94-142, the Education of All
Handicapped Children Act, and 7 percent for ESEA Title
IVC, innovative practices and curriculum. Much of the
evaluation money made available through federal programs
is controlled by the state or local program
administrators. This tends to put the evaluators in
competition with program administrators for resources.
Evaluation projects may be approved or disapproved on the
basis of their acceptability to the officials who run the
programs. Bernstein and Freeman (1975) suggest that it
is advantageous to have the program staff play a role in
the research process, preferably by having both the
program and the evaluation units be part of the same
overall organization. But unless an evaluation unit can
operate with some independence within the overall
organization and is given direct access to the leadership
of the organization, it cannot (and will not) be trusted
to produce credible work.

Recommendation C-2. Congress should separate funding for
evaluations conducted at the state and local levels from
program and administrative funds.

   The first reason for this recommendation is that such
a separation will allow greater accountability for how
evaluation money is being spent and who spends it. The
current arrangement for most programs is to have
evaluation money come from local program funds or from
state administrative funds. No separate accounting is
necessary. This makes it impossible to know how much of
the federal money potentially available for evaluation is
actually used for that purpose at the state and local
levels. It is therefore impossible to judge whether
inadequate performance of specified evaluation tasks
comes about through lack of funds, inadequate training,
or other factors.
   The second reason for the separation is to introduce
greater integrity to state and local evaluations. Under

1̇ ̇ ̇

present circumstances, whatever amount of money is
invested in evaluation is, in too many instances,
controlled by those who administer and run programs.
This puts the quality and credibility of evaluation
activities in jeopardy. As long as program
administrators control evaluation funds, resulting
evaluation activities will be suspect. If evaluation is
to be an independent function that can provide an outside
view of program operations and effects, it must be
separately funded.

As a specific way of accomplishing the separation,
Congress may wish to consider a required percentage
set-aside for each program that would be devoted to
evaluation activities at the state and at the local
levels, with due consideration of thresholds below which
no activity can be carried out adequately. Such a
set-aside provision should be accompanied by reporting
requirements that account for the money spent and that
summarize evaluation results and their application. Over
time, it will then be possible to judge whether the
investment in evaluation is yielding the desired results
in terms of program monitoring and improvement.

Capability

The competence and resources of the personnel charged
with evaluation responsibilities constrains their ability
to produce evaluations of acceptable quality. Only some
school districts, particularly the large urban or
suburban systems, have well-trained and sophisticated
evaluators. For many smaller agencies with limited
resources, staffing is inadequate for any of the complex
evaluation tasks such as process or impact assessments.
As Holley (Appendix C:258) notes:

> In most states certification standards are
> applied to personnel in federal programs. For
> example, a counselor, administrator, or supervisor
> must be certified to fill those roles in most
> states. In general, evaluators are not certified
> and no such standards are applied to the personnel
> filling the role of evaluator. In some LEAs and
> SEAs, the federal program director or coordinator
> may bear full responsibility for evaluation and
> even in agencies with substantial evaluation units,
> small federal evaluations may be completed by

101

program staff. Typically, where program staff are
given the responsibility for evaluation, they will
have neither training nor experience in evaluation
methodology, measurement, nor statistical
analysis. The author has observed many small
school districts in which the person charged with
Title I program evaluation is a reading teacher,
not only with no training in evaluation, but with a
weak background in mathematics.

Even when third-party evaluations are used, this does not
ensure either lack of bias or high quality, since school
personnel charged with selecting contractors may or may
not apply appropriate selection criteria. Moreover, the
competency of personnel in contracting organizations used
by local systems varies as much as that in the systems
themselves.

State agencies, in addition to carrying on their own
mandated and discretionary activities, are also charged
with a variety of responsibilities with respect to
evaluations carried out by local school systems.
Depending on the legislative provisions in a given
federal program, these may include "monitoring the
compliance of its districts with federal evaluation
guidelines, aggregating, analyzing and reporting data on
the state-wide impact of federal programs, and ensuring
that LEAs receive proper technical assistance in program
development and evaluation efforts" (Pion, Cordray, and
Boruch, Ch. 4:7 in Boruch and Cordray 1980). The size
and capability of evaluation staffs vary considerably
from state to state, and it is not necessarily
proportional to the school enrollment or to the number of
federal programs administered. Many states do not have
the capability to do more than minimally comply with
federal requirements, that is, forwarding the data
supplied by the local agencies.

Recommendation C-3. Congress should institute a
diversified strategy of evaluation at the state and local
levels that would impose minimum monitoring and
compliance requirements on all agencies receiving federal
funds, but allow only the most competent to carry out
complex evaluation tasks.

The Congress should require the Department of
Education to submit detailed program performance data.

Therefore, all state and local agencies receiving federal
funds for education programs should be required to
provide accounts of the allocation of program funds and
of program coverage. When specific services and
procedures are mandated, these too should be assessed for
compliance with the law (implementation accountability).
To accomplish this requirement, it may be necessary to
spell out in legislation dealing with evaluation
activities the resources, coverage or target groups, and
program services to be reported on by each recipient unit
(local education agency, state education agency,
community based organization, or other public or private
agencies). Congress should also require that the
Department institute quality control procedures that will
ensure usable and comparable data on program funding and
coverage.

Evaluation tasks that go beyond accountability
questions--for example, the assessment of educational
impact or the identification and testing of alternatives
that might lead to improved programs--should be a
selective activity rather than imposed on all, regardless
of competence and funds available. This recommendation
is not meant to suggest that creativity in providing
effective education cannot be found in school systems
with limited resources. Inventive teachers and
administrators have always found ways of applying the
lessons learned through experience to their classes and
their programs, but they do not do it through formalized
evaluation (David 1978). The task of understanding
promising approaches and applying such understanding to
program improvement at various sites is an extremely
ccmplex one that needs considerable investment of fiscal
resources and the skill of highly trained people who are
unlikely to be available to every school system and state
agency in the country. Nor is it necessary that every
site carry out that type of evaluation. If more were
known about how to provide effective services through
studies carried out at a limited number of sites and if
school systems were then encouraged to try those
alternatives that appeared most promising, program
improvement could be expected.

The description by Holley (Appendix C) of three
alternative means of funding local evaluations documents
the utility of providing discretionary funds on a
competitive basis for program improvement. Congress may
wish to consider authorizing a grants program for school
systems that would allow funding of the most promising

103

proposals for program improvement based on evaluation of program alternatives that appear to be effective in a given context (see Recommendation D-5 above).

Recommendation D-8. The Department of Education should explore alternative approaches to technical assistance for state and local evaluation needs.

The technical assistance needs of state and local agencies are not uniform. They vary with the size of the agency, the sophistication of the agency's evaluation staff, and with the complexity of the federal program activity in the agency. The regionally based technical assistance centers associated with Title I are one approach to meeting such needs. Whether the TACs are the best form of assistance for all agency types and sizes and whether the services they provide are adequate to all needs should be explored more extensively.[12] For example, the development of technical assistance capabilities in state agencies that also have authority and responsibility for supervising local activities might be a more reasonable and effective alternative. The National Institute of Education used such a strategy in building dissemination capacity within state agencies (Raizen 1979). Or the support of state, regional, or national networks of evaluators might permit the joint exploration of complex problems for which solutions do not yet exist (see Appendix C). Or seminars that bring together evaluation practitioners with representatives from a number of different disciplines could increase the awareness of alternative research techniques that might be brought to bear on complex problems and issues.

Technical assistance should also encompass organizational and personnel questions. Evaluators are often recruited and hired by people with little understanding of the skills required in the practice of evaluation. Personnel officers may, for example, be unaware of the types of degree they should require or of the types of candidates to interview. Consultants are hired to do evaluations, but their qualifications and training must frequently be reviewed by staff members unacquainted with evaluation. The relationship of evaluators or an evaluation unit to program administrators, executives of an education agency, its governing board, and public groups are often not carefully considered or are submerged in more powerful

carefully considered or are submerged in more powerful
organizational considerations. Technical assistance in
the area of evaluation organization and personnel
policies could draw on much work done already by some
state and local agencies as to optimal institutional
arrangements, personnel requirements, and procurement
policies for extramural work.

   In particular, state and local agencies need to be
aware of the desirability of separating the evaluation
unit from program administration. Especially in the case
of impact assessment, there is an obvious conflict of
both intellectual and monetary interest. Evaluators
should in general be outside evaluators, and evaluations
should not be controlled by the program administrators.
The case is more ambiguous for formative
evaluations--those that are aimed at improving programs.
Responsible program administrators should be doing this
kind of self-evaluation as a matter of course, but there
are also powerful advantages of having outsiders do this,
kind of evaluation: outsiders bring a fresh and unbiased
view and are likely to see new ways of solving problems
in program administration and new approaches for
improving program benefits. They are also not
constrained to cover up inadequate performance, as
internal evaluators may be inclined to do. The best
approach may be to encourage continuing in-house
evaluation efforts, but also to encourage agencies to
make greater use of qualified outside evaluators.
Technical assistance should help agencies organize their
evaluation activities in such ways that they can derive
the maximum benefit from their (and the federal)
investment in this area.

Recommendation C-4. Congress should require an annual
report from the Department of Education on all evaluation
expenditures and activities, including those at the state
and local levels.

   The current evaluation report delivered to the
Congress annually should be expanded to cover all the
evaluation activities within the Department as well as
those carried out by state and local agencies with
federal education funds. Past annual reports have
concentrated on the activities of the central evaluation
unit; they have not been comprehensive with respect to
evaluation activities carried out elsewhere in the

1 ⅋

Department. More importantly, no analogous report is now required of evaluation activities carried out at the local and state levels; even figures on federal dollars spent on evaluation at these levels are unobtainable, let alone any substantive account of either mandated or discretionary activities. It is therefore impossible to discern to what effect evaluation dollars are used at these levels except through special studies. Until more complete accounts are available of the total extent and nature of the activities carried out, the quality and management of evaluation cannot be improved.

The Department's report should specify the amounts of federal dollars spent for evaluation at the national, state, and local levels, and breakdowns of funding should be given by type of activity. Summaries of studies under way, findings and critiques related to completed studies, and their application to improvement of the substance and management of programs should also be included in the report. In addition, Congress may wish to request a brief report or special section on "What Has Been Learned," which draws from all relevant sources of knowledge--including evaluation and research not supported through federal education funds--to consider how programs can be made more effective through changes in legislation, management, or program strategy.

NOTES

1 The cited studies cover several social service fields. Evaluations in education may in fact have a better record than some others. Rezmovic (1979), in summarizing reviews of evaluation studies in the criminal justice field, finds that there are very few if any studies without serious shortcomings that jeopardize the credibility of study results. She cites Logan (1972), who examined 100 correctional research studies and found not one that met minimal methodological requirements for testing effectiveness.

2 We use Sharp's definition (see Appendix B) of "private-sector performers": all those not connected with a university or with a public education agency, local or state."

3 Major performers are defined as those that spend $1 million or more on educational research and development.

4 The term 8-A refers to a special form of
noncompetitive awards. An 8-A firm is a small
for-profit business concern that is owned, controlled,
and operated by one or more person(s) who are socially
and economically disadvantaged. To be eligible, the
concern must have submitted a business development
plan to the Small Business Administration (SBA), which
must have approved it for SBA assistance. An 8-A firm
can be selected to deliver goods or services to the
federal government without having to compete with
other firms.

5 A resource list compiled by NIE of minority firms
competent to do R&D work in education during that
period contained 185 entries; about two-thirds were
8-A certified.

6 Some of these programs are the Graduate and
Professional Opportunities Program (GPOP) in the
Department of Education, the Minority Fellowship
Programs in NIMH, the Minority Postdoctoral Fellowship
Program and the Women and Minorities Program in NIE,
the Minority Access to Research Careers (MARC) in both
NIH and ADAMHA, the Minority Fellowship Program in
NSF, and the Health Center Opportunities Grants (HCOG)
in HRA.

7 This information, including the dates given in Table
1, was provided by Priscilla (Pat) E. Dever,
Administrative Officer, Office of Program Evaluation,
U.S. Department of Education. We are grateful for her
help and patience in responding to our inquiries.

8 The 16 regional educational laboratories and R&D
centers have a special relationship with the federal
government through which they receive core funding
outside the competitive process, some of it for
evaluation studies, though they may--and several
do--also bid on RFPs. Of ten $5-million-plus
performers of educational R&D, two are regional
laboratories; nearly all these institutions fall into
the $1-million and over or "major performer"
category. Because they have long-term relationships
with the Department, they are in a favorable position
to receive contracts for evaluation work.

9 This provision was enacted at the behest of state
education agencies so that they could plan adequately
for their own data collection systems. It is
questionable, however, whether evaluation studies that
gather one-time information (even if collected more
than once, as in pre- and post-testing or in

longitudinal studies) and are usually done on a sample
basis would affect these data systems to any extent.

10 A cooperative agreement is a type of award used as an
alternative to a contract when a project requires
substantial involvement of the sponsoring federal
agency during project performance. "Substantial
involvement may be necessary because the project is
technically or managerially complex or requires close
coordination with other federally sponsored work.
Examples are policy studies, projects requiring
complex subcontracting, large curriculum projects, and
evaluations of federal programs. For a detailed
definition, see P.L. 95-224.

11 A basic ordering agreement is a written instrument of
understandings between the government and a contractor
that sets forth negotiated clauses to be applicable in
future contracts, including a description of supplies
or services to be furnished and of the method for
determining fees to be paid. This instrument is
generally used in conjunction with a selected group of
contractors found to be qualified to furnish the
specified supplies or services when needed.

12 A recent evaluation of the TACs (HOPE Associates
1979:60) found diverse views of their effectiveness
among state agency personnel. One of the reviewing
panel's recommendations was that

    . . . the Office of Education begin to
    investigate, during the period of the next
    contracts for Technical Assistance Centers,
    the possibility of a future system that has
    flexibility to accommodate to: the diversity
    of state and local capabilities and needs,
    and also the enlarged objectives of Title I
    evaluation technical assistance, particularly
    including the uses of evaluation for local
    program improvement and the strengthening of
    local evaluation capacity.

4
Using Evaluation Results

A frequently voiced statement about evaluation is that
evaluation findings are rarely used. Often this type of
statement is followed by the criticism that few policies
have been changed and few programs either terminated or
started because of the findings from evaluation.
Implicit in this criticism is a belief that "utilization"
means direct and often immediate incorporation into
policy and program. The criticism carries weight mainly
for those who have a definition of utilization that comes
close to making it a substitute for the political
process. We do not take that position. In our view,
utilization takes on a variety of forms, not all of them
immediately evident.

   Indeed, we maintain that the main goal that evaluation
can rightfully espouse is that of being "useful": that
is, evaluation-based knowledge is disseminated to those
audiences that have a need or an interest in it, is
presented in a fashion that is understandable to them,
and is addressed to the policy questions that are
relevant to them. Evaluation cannot and should not
substitute for the political process. Nor can evaluators
ensure that evaluations are used. The best one can do is
to make sure that evaluation findings are available to
those who might want them and that the findings address
the issues of concern in an understandable and
responsible way.

   Because much of the difficulty with utilization
centers around the differing meanings of that term, in
the first two sections of this chapter we discuss the
varieties of utilization and some of the limitations that
constrain the use of evaluation findings. Next we

summarize the evidence on how evaluations actually are
used and show that considerable use is made of evaluation
results, even though evaluations rarely shape social
policies in a sharp and immediately obvious manner. The
next section discusses the research literature on how
science-based knowledge is used and how its use can be
enhanced; the final section identifies the various
audiences for evaluation findings, their information
needs, and what the Department might do to better serve
those needs.

DEFINING UTILIZATION

"Utilization" has been used to cover a variety of things,
a semantic imprecision that lies at the root of a common
impression that evaluation results are rarely
"utilized." One major source of difficulty lies in the
failure to distinguish between dissemination and
utilization. Another major source of difficulty is that
"utilization" has been used to mean overt changes in
social policy and programs as well as uses of evaluation
findings that fall far short of changing social policy.

Dissemination and Utilization

It has been recognized for some time that dissemination
of knowledge does not necessarily lead to its use, though
it is a requisite first step.[1] For purposes of this
report, dissemination of evaluation findings means the
deliberate communication of knowledge derived from
evaluation activities; utilization refers to the use of
such knowledge when decisions are made about educational
policies and programs. Such use may include instituting
a change as a result of having considered the evaluation-
based knowledge. However, "dissemination" is often used
to mean or imply utilization and subsequent change; that
is, utilization and change are viewed as an almost
automatic by-product of communication. This use of
"dissemination" is unfortunate and misleading because
recent empirical studies on utilization and change make
it clear that knowledge, however packaged and
disseminated, has little compelling power in its own
right (see, for example, Caplan et al. 1975, Caplan 1980,
Berman and McLaughlin 1975-78, Human Interaction Research
Institute 1976). These findings hold for

110

purpose-specific information such as program evaluations
as well as for forms of knowledge for which the relation
between knowledge production and intended use and
audience is less obvious.

The distinctions between dissemination, utilization,
and change are important to keep in mind. Dissemination,
because it is largely under the control of evaluators and
sponsors, can be improved by self-conscious efforts.
Improvements in dissemination strategies can usually be
made that, other things being equal, ought to lead to
greater utilization and to change when indicated. But
other things are generally not equal: the forces and
events impinging on decisions about programs may be more
powerful than evidence from evaluation activities.
Moreover, such evidence is often couched in statistical
terms that are not translated into terms having
substantive meaning or that may not be substantively
significant.[2] Steps can be taken to ensure wide and
effective spread of information and thereby improve the
likelihood of utilization, but we know of no means that
can ensure utilization, let alone change.

Forms of Utilization

There is currently a very strong emphasis on using the
results of evaluation for making specific decisions at a
given time; for example, when legislative or budgetary
decisions are anticipated or when changes in program
regulation or management are being considered.
Sometimes, this perspective is appropriate, as was the
case for the NIE compensatory education study, which
began with some specific issues and fairly well-defined
problems (National Institute of Education 1976) and chose
to investigate factors that could be controlled through
changes in policy (Hill 1980). The desire of those who
initiate and pay for evaluations (Congress, the
Department, state and local governments) to obtain
immediately applicable results is understandable, but it
can lead to inappropriate expectations.

In particular, the grounds for decisions cannot always
be specified beforehand. For example, funding decisions
are sometimes declared to be the policy questions that
the results of evaluations are to address. Yet funding
decisions are generally made on a variety of grounds,
many of which cannot be addressed by evaluations, as has
been amply demonstrated by the history of impact aid,

111

Head Start, Follow Through, bilingual education and other
programs that became popular with beneficiaries and
service deliverers. A program may develop such strong
constituencies that the results of evaluations become
largely irrelevant to funding decisions. As another
example, the evaluation of alternative compensatory
education interventions used in Follow Through was to
identify the most effective model for wide-scale
implementation (Elmore 1975). It turned out, however,
that there was more variation within models than between
models; moreover, increased funding to permit increases
in the program never materialized.

The possible decision issues also change over time in
unpredictable ways. Turnover among federal executives is
high.[3] Questions that are tied to the perspectives of
an individual decision maker or of a particular
administration may no longer be of interest when a new
executive or administration takes over. Decisions also
change as educational priorities change over two or three
years, even under the same administration.

In short, while evaluation for specific decisions
appears to be a sensible strategy to follow, such a
strategy may be much wasted effort. The issues involved
in a decision that is to be taken at some time in the
future are not easily predicted. Hence an evaluation
started today that is directed towards the specific
decisions envisaged two years hence is just as likely as
not to miss the mark because the issues in the decisions
will have changed.

One implication of the above is that evaluations
should seek out questions of lasting significance and
provide knowledge that can be used and reused, knowledge
that may be exploited in several different ways over time
in addition to furnishing short-term information
(Chelimsky 1977). Involved here are differences in types
of knowledge application, i.e., knowledge for
understanding versus knowledge for immediate action,
sometimes also referred to as conceptual use (indirect
impact on decision perspectives) versus instrumental use
(direct, mechanical application) (Weiss 1977). To ensure
the maximum utility of any major evaluation, it should
address questions appropriate to both uses. Adopting
this principle has consequences for the planning of
evaluations (see Recommendation D-10, below).

A third use of evaluation can be called
legitimization: the primary purpose of the evaluation is
something other than to develop knowledge about a

program. The reason for initiating the study may be more
important than the eventual results, such as meeting
legal requirements for evaluation, demonstrating the
objectivity of an agency's decision making, or supporting
some particular point of view (e.g., the need for more
program funds). Though such motives are not often
overtly acknowledged, the use of information that results
from such evaluation studies is not necessarily
illegitimate provided valid data are reported and
interpreted honestly.

Misuse and Deliberate Nonuse

One of the problems in defining the process of
utilization is that not all study results ought to be
used and that deliberate rejection or nonuse of results
that are faulty or otherwise inapplicable is preferable
to misuse. Misapplication of results is as much a
negative consequence of evaluation as lack of
application, and deliberate nonuse may represent rational
decision making as much as does appropriate
application.[1] The problem is that the deliberate
nonuse after results have been carefully considered and
dismissed for valid reasons is difficult to distinguish
from the failure to use evaluation results for other
reasons.

Aside from nonuse for valid reasons, it is important
to distinguish between the misuse or nonuse that results
from of lack of judgment and that which has as its
motivation the suppression of valid information. Persons
who may not be fully aware of the standards of quality
that should be applied to evaluation studies may hail the
results of faulted work and condemn on seemingly
technical grounds quite well-executed studies. This lack
of judgment calls for attempts to inform potential users
of the standards by which various types of studies should
be judged. The recommendations made elsewhere in this
report on open and systematic review of evaluation
studies should be helpful in judging quality. (Our
recommendations on training in Chapters 3 and 5 are also
intended to address this problem.)

Deliberate misuse or nonuse of evaluation studies is
in many ways more difficult to deal with. First, it is
difficult to detect motives. Second, it is not likely
that persons deliberately abusing evaluation studies
would be likely to be dissuaded by arguments based on

considerations of quality. The best that evaluators and
the Department can do is to make sure through re iew of
evaluations that those that are defective are clearly
identified and that exemplary evaluations are also
clearly identified. Full publicity should be given to
the evaluation review procedure and its results.

LIMITATIONS THAT CONSTRAIN USE

Just as the definitions related to utilization are
important to understand if one wants to improve the
utilization process, so are the functions of knowledge
within any agency or for individual decision makers, at
whatever level.[5] Evaluation cannot and should not
replace the political process. This means that an
automatic translation of evaluation findings into policy
decisions is neither desirable nor to be expected.
Policy makers cannot override the ideological, political,
and financial limits they face, though these limits are
themselves subject to change over time, aided by the
accumulation of knowledge. Decision makers and managers
are not always able to take actions that seem to the
researcher the "best" form of intervention or
implementation. Both the feasibility and the
acceptability of a change in public policy are as
critical as science-based knowledge in determining the
course of a decision (Ezrahi 1978). Thus a program that
is feasible and effective but likely to arouse the
resistance of significant constituencies, or that can be
funded only at the expense of some other more desirable
program, or that is liable to antagonize school
administrators or teachers, is not likely to be adopted.
Nor should it be, given that legislatures and public
officials are expected to be responsive to such
realities. There is no special democratic license given
to the results of evaluation that allows such results to
override the ordinary political considerations that
surround education just as they surround other important
areas of social policy.
    So it is important that, from the outset of any
evaluation, the range of options and political realities
regarding timing, variables, and likely decisions be made
clear by the likely users. Early collaboration between
researchers and decision makers in planning the research,
identifying variables, specifying time frames, and
defining the problem under study will help toward wiser

1 1

and more profitable use of social science research,
especially program evaluation, within the political
context of social probem solving (see Recommendations C-1
and D-1).

Though we use the term "decision maker" in this
report, we do not mean to imply that decisions about
programs are made as if there were sovereign rulers in
government. Yet evaluation reports are often written as
if such individuals existed and were able and ready to
act on evaluation findings and recommendations. As we
noted above, the persons who initially ordered and
collaborated in planning evaluations and their
utilization may have moved on to other responsibilities
by the time findings are available. Their successors
often have less interest in or less understanding of the
purpose of the evaluation. In addition, interests
sometimes shift rapidly at the top echelons of government.

Having some documentation of the purpose and
importance of a study that can be referred to after the
authority for decisions has changed would help in
utilization. However, as has become evident from
research on organizations (see, for example, Cohen and
March 1974, Cohen et al. 1972), policy is often not
"made"; rather, it accumulates by slow accretion. New
information may actually slow down the process since it
may make decisions more complicated. Thus, one has to
think of policy formulation and decision making as
involving different stages, different people, and a
process of absorbing and digesting all types of
information: tested empirical findings from evaluations
are only one of those types.

While the reduction of ignorance may always be
desirable, it is not synonymous with the reduction of
risk. In fact, new information may produce considerable
risks as it enters an organization. Perturbations go
through the organization--established assumptions and
ways of doing things become threatened, agenda priorities
and budget line items may be thrown into question, and so
forth. The common response to such threats is to let
procedure take precedence over substance and to ignore
the message of the new information in the interest of
preserving established procedures and structures. To the
outsider, it may appear that the information is ignored,
though it may be used informally. Studies carried out on
the use of knowledge among upper-level federal officials
in the United States and abroad show that the control of
information is more important than its use (Caplan

1980). The bureaucratic nature of state and local
educational agencies has been amply documented (Murphy
1974); maintenance of the organization is also a priority
goal. So, if knowledge use is to be furthered, stress
must be placed on understanding bureaucratic rationality
and on being nonjudgmental about it. It really is no
less "correct" than individual or scientific rationality,
but it is different and will deal differently with
information.

EVIDENCE ABOUT UTILIZATION

To what extent is the impression correct that evaluation
results in education are little used? Who does use
evaluation results and who does not? The most
comprehensive review addressing this topic consists of
the recent case studies done by Leviton and Boruch (Ch. 6
in Boruch and Cordray 1980) and the accompanying analysis
of the existing literature on evaluation utilization.
The analysis, which generally confirms the findings of
earlier research, is summarized below.

First, despite the difficulty of tracing utilization,
there are a number of well-documented cases both at the
national and at local levels in which evaluation findings
were used directly in modifying laws or regulations,
influencing choices of curricula or instructional
strategies, or altering management practice. For
example, of the 42 evaluation activities included in the
section on use in the 1979 Annual Evaluation Report (U.S.
Department of Health, Education, and Welfare 1979b),
one-third were specifically cited in congressional
documents or led to identifiable revisions in regulations
and other management procedures.

Second, cases of conceptual use, or contribution to
the accumulation of knowledge about a program, are
obviously more difficult to verify. Nevertheless, there
is evidence from interviews with congressional staff
(Florio 1980) and research on the behavior of federal
executives (Caplan et al. 1975) that some of the major
sources of information (e.g., the Congressional Reference
Service) used in Congress and by executive agencies are
based on research evidence, including evaluation
findings. Often, such research-based information is used
for framing issues, developing program ideas, and general
oversight rather than for immediate decision making.
This type of knowledge use is not always apparent even to
the user, let alone recognized by an outsider.

116

Third, in the last few years, the majority of
evaluation studies have been concerned with
implementation and managerial process--the type of study
most likely to lead to direct application. In this,
evaluation is not different from other social science
research; Caplan et al. (1975) found that more than half
the use of social-science-based knowledge by federal
executives was to increase administrative efficiency and
organizational control. The use of results from program
effect studies has been more difficult to discern, and
even when such studies are cited, it is not the findings
on effects, but those on coverage and management that are
used. The evaluation study of the bilingual education
program provides a good example (Danoff 1978).

Fourth, a continuing problem in relation to
utilization is the failure to spell out the ways in which
the information developed by a study could be applied.
What policy options appear preferable to reach certain
goals? What management strategies deliver services
effectively? What are the outcomes of different
curricula in different types of classrooms, for different
types of students? When evaluation studies address
questions not perceived as important by a particular
audience, they are likely to consider the results
irrelevant and useless. For example, a number of local
sites have reported that the data required by the federal
government on Title I and other education programs are
not useful to the local agency (David 1978), while others
consider such data useful but needing to be augmented by
specific local studies in order to gauge program progress
(Boruch, Leviton, Cordray, and Pion, Ch. 6 in Boruch and
Cordray 1980).

Fifth, there has been little attempt to specifically
reach audiences concerned with equal educational
opportunity. Women, minorities, and handicapped people
generally believe they have limited access to social
science research and evaluation processes that they see
as affecting programs that are significant to them.
Because of this perception of exclusion, some of the
largest groups involved in equal opportunity issues, such
as the NAACP, ASPIRA, COSSMHO, the National Urban League,
and the National Council of La Raza, are developing their
own capability for research and development or have begun
to work closely with research organizations willing and
capable of addressing issues of interest to minority
groups. The Council for Exceptional Children performs a
similar function for programs serving handicapped

117

children, as do women's organizations for programs of
concern to them. As long as groups representing
beneficiary interests see themselves as peripheral to the
sharing of information produced by evaluation, there is
likely to be unnecessary controversy and friction.[6]

TOWARD INCREASED UTILIZATION

The preceding sections have a tempted to define various
types of knowledge use, discussed the setting or context
for use, and briefly reviewed the evidence on the degree
of use. Before considering what might be done to
increase the use of evaluation results, we summarize what
has been learned about the utilization of research
knowledge in general. The research literature is replete
with recommendations on how to improve the likelihood
that knowledge will get transferred from producer to user
and actually used (see, for example, Havelock 1969, Davis
1973, Glaser 1973, Havelock and Lingwood 1973, Rogers and
Shoemaker 1971, Zaltman et al. 1973). Those
recommendations tend to cluster around two sets of
factors: the nature of the information and how it is
communicated.

Nature of the Information

The ways in which knowledge is produced and is perceived
by its potential audience(s) affect its use. The
important characteristics of knowledge associated with
increased likelihood of use can be summarized as
intuitive correctness, objectivity, and relevance (Caplan
1977). Obviously, there is not much that researchers can
do to produce knowledge that fits the first
characteristic, that seems to match common sense or to
"feel right." However, intuitive correctness is probably
most important only in the early stages of policy
formulation, for needs assessment and for considering
intervention possibilities. Perceptions of objectivity
are usually enhanced by distancing evaluation from
program operations, but, as noted in Chapter 2, this may
also make results less relevant for some audiences. The
reverse is true as well.[7] Relevance involves
continuous interaction between the primary audience and
the researcher, although that may affect the researcher's
objectivity.

118

There are several important elements in achieving relevance:

● Negotiated content. Evaluators, sponsors, individuals, and groups comprising the primary audience(s) (if other than the sponsor) and action sites or program managers must negotiate what issues and information needs can be addressed in terms of researchable questions and what types of data it will be possible to collect at program sites. Such negotiation is not a one-time-only task; it should proceed throughout the evaluation so that the study is not stymied or does not turn out to be irrelevant.

● Appropriate research forms. Insofar as methodological limitations allow, the research should aim to use the policy maker's or primary user's definition of the problem. Researchers too often tend to define the research to fit methodologies rather than the interests of the likely audience. The law of instruments has a way of taking over: that which can be measured is measured, whether or not it addresses objectives or concerns of interest to the policy makers or program managers.

● Realism. The research questions addressed and the interpretation of results must deal with options that are realistic for the decision makers expected to take action. The variables under study should be ones that are politically malleable: that is, they can be changed, if necessary, in order to improve policy or program substance. For example, periods of reading instruction can be lengthened, but a 1:1 student/teacher ratio, even if effective in teaching reading, is unrealistic on a wide scale because of its cost. Implications and recommendations must take into account the constraints of likely users, such as political acceptability or budget limitations.

● Timeliness. It is especially critical for direct knowledge application that information be timely. If a study is to provide input to legislative or funding decisions, but is not geared to the authorization calendar or the budget cycle, it will be irrelevant to the primary audience(s). While what may be relevant today may not be relevant tomorrow, increased contact among parties at interest and evaluators will improve the probability that relevant questions will be addressed.

Attention to these elements was a major factor in the success of the NIE compensatory education study (Hill

1980). And portions of effectiveness studies deemed
relevant, namely those having to do with coverage and
resource allocations, have been used in formulating
legislative amendments, appropriations, and changes in
regulation, even when the findings on effects appeared to
be ignored: for example, the histories and use made of
the sustaining effects study (Systems Development
Corporation 1976) and the Title VII bilingual education
study (Danoff 1978). (Citations in congressional
documents of these studies and other documented uses are
given in Boruch, Leviton, Cordray, and Pion, Ch. 6 in
Boruch and Cordray (1980).) In Chapter 5 on the
organization and management of evaluation activities, we
make some recommendations pertinent to increasing the
relevance of evaluation studies. Timeliness in
particular and current impediments to completing studies
on time are treated at some length in Chapter 5 (and also
in Chapter 3). We reiterate the need for quick-response
evaluation capability on part of the Department, as well
as sophisticated planning of major evaluation tasks that
will yield at least some useful results at the time they
are needed by primary decision makers in Congress or at
the top levels of the Department.

Communication of the Information

The many factors that have been identified in the
literature as enhancing the transfer of knowledge and its
use can be grouped under two headings: communicability
and linkage.[8] Communicability encompasses matching the
style of communication used by the researcher or other
transfer agent. (see below) to that of the primary
audience(s). Since researchers are not necessarily the
most effective communicators, nor will they always be on
call when needed, linkage by means of transfer agents is
necessary.
Several principles about communicability have emerged
from the literature and successful practice:

• Intelligible reports. Reports to primary
audiences should be tailored as much as possible to their
needs and their situation (Patton et al. 1977). Language
should be understandable and situationally applicable;
e.g., papers and reports written for scholary audiences
are rarely appropriate for the primary or other
audiences. Too often, social science researchers write

for their colleagues and, even when studying issues of
pressing public concern, tend to emphasize the esoteric,
counterintuitive, or paradoxical. Social scientists in
the United States have a special fascination for numbers,
but more emphasis should be given to the substantive
meaning of evaluation findings, not to their numerical
properties and the niceties of the statistical analyses.
Reports should avoid jargon, be written in plain English,
and address in a straightforward manner the issues
relevant to the intended users and their informational
needs. If a number of different audiences have primary
interests, several versions (or translations) of a report
may be necessary.

• Accentuating the positive. Whenever possible,
recommendations ought to highlight positive action steps
that can be taken. Things not to do are important to
recognize as well, but they rarely carry the same kind of
reward for individuals in a position to act.

• Live communication. The print medium is not the
only nor even the most effective means of communication.
Face-to-face interaction and reporting through
conferences provide alternative mechanisms. This allows
clarifying questions and making sure that the most
important points are covered. Information is more likely
to be used when it comes from sources that are trusted,
and human beings trust other human beings whom they have
found to be reliable in the past more than they trust a
computer terminal. Redundancy of communication has
proven effective, so that optimal dissemination
strategies are likely to include both oral and written
communication.

As we noted above, linkage is the term used to cover
the gap that may exist between researchers and the
audiences for their findings. Techniques to create
linkage derive from research on communication and the
spread of innovation (Katz and Lazarsfeld 1955, Rogers
1962). Lippitt (1965) and Havelock and Lingwood (1973)
single it out, as the most critical step. The issue is
not just mechanisms of knowledge transfer, but
information management, storage, retrieval, and knowledge
synthesis. Past RD&D (research, development, and
diffusion) efforts by the Office of Education were
premised on the assumption that knowledge transfer and
linkage through organizational arrangements would be
effective, but the example of the Congressional Reference
Service shows the importance of people who act as the

translators or linkage agents. Experience with the
Educational Resources Information Center (ERIC) also
indicates that a computerized system for storing and
retrieving research information works best when a live
person acts as intermediary between the questioner and
the system. Linkage can be performed by in-house staff
(for example, individuals in the evaluation unit or in a
separate dissemination component) or by parties external
to either the research or the user communities.
   Some important factors that affect linkage include:

   • Responsiveness to differences. Transfer agents
or groups must be responsive to differences between
researcher and audience and to differences among
audiences--perspectives, values, motivation, and
language. They must know how to translate from one to
the other and when direct interaction should take place
and when not. (For example, some researchers make
excellent congressional witnesses, others--equally
eminent in their field--do not.)
   • Mediating problem definitions. Even at the
beginning and during the course of a study, transfer
agents can be useful because--speaking the language of
both the researcher and the audience--they can help
define policy decision problems in researchable terms.
This role can be especially important when the intended
user is not the immediate sponsor of the evaluation and
therefore does not have automatic contact with the
researcher. Problem definitions and criteria used by
those requesting an evaluation must be understood by the
researcher and be a guide to what will be done in a
study. They must also be clarified so as to be
researchable, or the reasons they are not must be
conveyed to those requesting the evaluation. (As we
noted in Chapter 2, examples of unresearchable problems
are the measurement of effects for diffuse or broad-aim
programs for which objectives cannot be specified, the
measurement of the aggregate effects of a program that
takes different forms in thousands of different locales,
or the effects of weak treatments administered in complex
settings.)
   • Human agents. Linkage is best achieved by people
rather than by cold-terminal (computerized) systems,
although this may change as the computer culture becomes
more pervasive and terminals become more accessible in
location and in language. At present, however, decision
makers are still used to face-to-face communication for

most important transactions, which only later get
codified in print (Chelimsky 1977).

• Open systems. Bureaucracies, including
legislative and executive agencies at the federal, state,
and local levels that deal with education, tend to be
self-referential systems; that is, people in
bureaucracies look for information that comes from the
inside and find it more credible. This characteristic is
also true of other people in the evaluation process, such
as the various interest groups. For example, teachers
tend to consult other teachers and their professional
associations when they need information; groups
representing minority interests have set up their own
research components. It also applies to knowledge
producers, i.e., researchers, particularly those who are
university-based and are not dependent for their
livelihood on communicating with potential sponsors of
evaluations. Transfer agents can help make all these
groups more aware of outside information. But to go
beyond awareness and expect linking or transfer agents to
increase responsiveness to information would require them
to understand the function of information in each group
and the risks that the use of information entails for
each.[9] Transfer agents are not likely to be able to
counteract behavior based on maintaining cherished
assumptions or well-established procedures and that
therefore has a need to ignore perturbing research
findings.

Recommendation D-9. The Department of Education should
test various mechanisms for providing linkage between
evaluators and potential users.

The Department might consider establishing a unit
charged with studying, developing, and instituting
knowledge transfer mechanisms and evaluating their
effectiveness. Alternatively, outside experts might be
charged with this responsibility. Appropriate activities
of a linkage unit, whether within or outside the
Department, would include:

• Helping assess proposed dissemination plans for
evaluation studies and suggesting improvements;
• Performing needed translations of evaluation
reports so that they can be understood by the intended
audiences;

Funding research (in conjunction with the NIE
dissemination research unit) on the access, transfer,
communication, and utilization of evaluation information
issuing from studies sponsored by the Department and
elsewhere;
    ● Developing effective techniques for the
synthesis, storage, and retrieval of evaluation studies
on a continuing basis; and
    ● Developing and installing regular procedures and
institutionalized arrangements designed to facilitate the
use of evaluation data on a day-to-day basis, at least
within the Department.

AUDIENCES FOR EVALUATION FINDINGS

If the main purpose of evaluations is to help develop
more effective policy and improve education programs, who
are the audiences that are likely to use evaluation
results in this way? What kinds of information do they
need? And how can evaluation planning be improved to
better serve those needs?
    Conventionally, evaluations at the national level have
been considered relevant to two primary audiences:
policy makers in Congress and in the federal agency
(i.e., the Department of Education) and federal program
managers. In this simple view, policy makers would use
the findings from evaluations to determine present and
future program needs and directions, and managers would
have a tool by which to improve the delivery of services
mandated in programs. As evaluation results have become
visible, however, it turned out that they have also
served as ammunition for critics of controversial
programs or as support for a program's advocates.
Federal legislators, convinced of the importance of local
decision making in education, have also been concerned
with local use of evaluation results to improve programs
within the local school system.
    Empirical evidence from studies of the use of
evaluations (e.g., Boruch, Leviton, Cordray, and Pion,
Ch. 6 in Boruch and Cordray 1980, Brickell 1974, Alkin et
al. 1979) has shown that not all of those audiences can
be served by any single overall study. The information
needs of diverse audiences with varying and sometimes
conflicting interests and perspectives make it virtually
impossible for one evaluation study to satisfy them all.
Policy makers may be mainly interested in coverage

issues, program managers in efficient delivery, and
recipients in issued of equal educational opportunity.
Each of these interests requires a different approach,
even different data collection.

Perhaps the clearest example of the problems of
diverse interests is the case of Title I evaluations
(Wisler and Anderson 1979, Cross 1979, David 1978, 1980,
Reisner 1980). The major evaluation strategy used since
the inception of this program has been collection of data
at the local level that, through aggregation at the state
and national levels, was to serve the information needs
of all three levels of government. The result has been
the generation of large quantities of data that have not
been useful at either the local or the national level--a
costly and frustrating process leaving all parties
dissatisfied. The failure of Title I evaluations has
been blamed on the lack of competence at the local level
to collect data that can be aggregated. While the
competence of some local evaluation units may be an
issue, the history of Title I evaluations illustrates a
much deeper problem, namely, the confusion of evaluation
purposes. The original intent of the congressionally
mandated local evaluations was to serve the needs of a
local audience, defined by some to be the parents of poor
children and by others to be the local school
administrators and teachers. Later demands for assessing
the overall effects of Title I spawned a complicated
system of aggregating from the local to the state level
and from the state to the national level. When it turned
out that data emanating from thousands of different
sources proved noncomparable, Congress mandated technical
assistance to the local systems to help with procedures,
designs, measures used, and problems encountered at the
local level. Models for evaluation designs were
developed and the technical assistance centers were
created to instruct local evaluators in proper use of the
models. Yearly costs for this assistance system now
stand at $12 million, more than half the budget of the
central evaluation unit. And yet complaints about the
utility of Title I evaluation information continue.
Local school systems find the data they are required to
collect by federal directive of little use to them and,
if they have the resources and the competence, they
conduct their own program improvement studies. At the
national level, Congress has consistently expressed its
dissatisfaction with the information it receives, as
evidenced by the rewriting of the evaluation requirements

for Title I that has occurred in every reauthorization of the program. Congress finally resorted to commissioning its own study, which was carried out by the National Institute of Education, a unit that was independent of the Office of Education (P.L. 93-380, Section 821). Leviton and Boruch (Ch. 6 in Boruch and Cordray 1980) summarize the evidence on the usefulness of the NIE study to its audience, citing specific changes in law and regulations in six major program areas directly traceable to study findings. Much of the success of this study as contrasted to all the other Title I evaluations is explained by its director (Hill 1980) as due to the extensive consultation with the primary audience, Congress.

To increase the probability that results will be used, the plans for an evaluation should spell out who the primary audiences are likely to be and how it is planned to reach them, so that both the substantive issues and the dissemination strategies can be negotiated with them. However, there will often be a number of secondary audiences. For example, an evaluation concerned with testing alternative curricula in career education to facilitate local choice may also affect the regulations governing federally supported vocational education programs. For evaluations conducted at the national level, decision makers (within the agency and Congress) and managers at the federal level are likely to take precedence. But where federal funds are made available for state and local evaluations, needs at those levels should be served.[10]

The Role of Planning

Although planning does not necessarily lead to an agenda that is subsequently carried out in detail, the act of planning always leads to an improved sense of priorities, provides a forum in which competing interests can reach accommodations, and induces an active as opposed to a reactive stance toward essential activities.

Recommendation D-10. The Department of Education should
institute a flexible planning system for evaluations of
federal education programs. (See Recommendation D-1.)

A flexible and workable planning system must have
several attributes. First, it ought to provide for
appropriate information for the predictably recurring
legislative cycles on education programs. This entails a
standard sequence of studies--timed to be available for
reauthorization and appropriation hearings--that will
furnish information on the coverage of programs,
descriptions of how they are run, and a synthesis of
information available at any given time of what can be
said about their effects. Second, there must be an
ongoing program of evaluation studies carried out at the
deliberative pace required to adddress problems that are
poorly understood. Third, the Department must have the
ability to respond to interesting questions that arise as
a result of ongoing research, changes in policy, or
development of new programs.

In the past, the central evaluation unit of the
Department has concentrated resources on massive studies,
in part because such studies require fewer procurement
actions to allocate available funds. But big studies
invariably take longer than anticipated and become highly
inflexible; hence they often end up addressing matters of
tangential interest to the audience at hand when they are
finally completed. Any evaluation plan for a major
education program should contain a series of linked
studies, some of which furnish factual information that
can be obtained in reasonably short time and some of
which address issues of long-term interest. Thus, at any
particular time and especially at predictably recurring
decision stages, one or more additional sets of findings
about a program will be available. Additionally, the
value of the whole evaluation plan does not depend on the
success or failure of a single massive study or on the
performance of a single contractor; there will always be
some useful studies resulting from the overall plan, even
though some may not turn out as hoped. In addition to
the plan for the NIE study of Title I, examples of such
evaluation planning are the original plan to evaluate the
Education for All Handicapped Children Act (U.S.
Department of Health, Education, and Welfare n.d.) and
the Department's new evaluation plan for Title I of ESEA
developed in 1979 (U.S. Department of Health, Education,
and Welfare 1979c).[11] The Committee applauds the

Department's direction in this respect and believes that
it will help make the Department's studies more relevant
to the immediate concerns of decision makers and
departmental managers. Before any costly evaluation
study is undertaken, however, ways in which it can inform
decisions and the risks of the evaluation questions
changing during the course of the study should be
outlined through the type of evaluability assessment
described in Chapter 2 or through some similar process.

The absence of a reasonable planning system in the
Department has had two deleterious consequences.[12]
First, it has given rise to an emphasis on activities for
"putting out the fire"--projects done in response to an
immediate crisis because no suitable information was at
hand when the question arose. Not infrequently, such
projects are irrelevant by the time they are completed,
either because the crisis has subsided or a different one
has arisen and attention has shifted. The emphasis on
addressing immediate concerns has reduced the
Department's ability to evaluate programs on a recurrent
basis in a fashion that would cumulate evidence on their
implementation and effectiveness over time. Studies to
develop and test out more effective program alternatives
receive even shorter shrift.

The second effect of the absence of appropriate
planning has been to create yearly uncertainty, beyond
that created by the budget process, about what studies
the Department will undertake. When yearly planning is
not set in the context of approved ongoing plans, the
approval process takes longer than necessary and may be
subject to capricious and arbitrary decisions. The
history of fiscal 1980, when it took 6-9 months to obtain
approval for initiating a study, provides a vivid example.

Recommendation D-11. The Department of Education should
establish a quick-response capability to address critical
but unanticipated evaluation questions.

No matter how flexible the planning system, there will
be a continuing need to respond quickly (within a 2- to
6-month time frame) to evaluation-related questions that
come from the Congress or from top-level Department
officials. Department staff charged with evaluation
responsibilities must be in a position to deal with such
requests. In some areas, in-house expertise may exist,
but even under the best of circumstances such expertise

cannot be expected to cover the great variety of topics
that may surface at various times. Several extramural
mechanisms are available for a quick-response capability:

 • Lists of contractors can be maintained who, as a
result of being found qualified in specified areas
through the RFQ process, can be awarded small contracts
within days for work that is limited in scope and time.
This mechanism in the form of basic ordering agreements
has been used by the Assistant Secretary for Planning and
Evaluation (ASPE) in the former HEW; the dollar limit on
contracts was $60,000.
 • Highly qualified selected organizations can be
awarded contracts that pay for a given number of
person-hours of effort, with tasks to be specified as the
need arises. This mechanism has been used in the
Department of Labor, with the limit for any one-year
contract set at $200,000.
 • 8-A contracts and awards to SBA-eligible firms
can usually be executed more quickly than other types of
contracts.

In order to be fully responsive to the information needs
of its primary audiences, the Department must be able to
combine a deliberative planning process that allows time
for field and constituency involvement with a
quick-response capability that can address unanticipated
but critical evaluation questions as they arise.
    The need to serve short-term information requests can
be considerably enhanced in any program by the
development of good management information systems.
Thus, for example, if a good management information
system had been in place, it should have been possible
for the Spanish/English bilingual education program
(Title VII) to have provided Congress with detailed
information on the ethnicity and language status of the
students being served. Instead, a study intended to
assess the impact of the program had to use a
considerable share of its resources for documenting
program coverage (Danoff 1978). Similarly, such
questions as the trends in composition over time of
students enrolled in education courses in colleges and
universities ought to be routinely collected as useful
and necessary background data on the future supply (over
or under) of teachers.
    For many programs that are not funded through the
Department, the provision of such management information

may be difficult to the point of impossibility. But for
federal programs, the Department should consider the
possibility that good management information systems may
provide much of the information that may be required
about a program for many decision-making purposes. Such
systems must be carefully designed, however, to provide
information that is likely to be useful, rather than
trying to cover all contingencies. As we note in Chapter
5 below, grantee reports have too often been collected
without ever being reviewed.

AUDIENCES FOR EVALUATION FINDINGS

The discussion of different audiences for evaluation
results that follows tries to indicate different
information needs for each. Two facts should be noted:
there are important distinctions within broad classes of
potential users or audiences, and sponsors are sometimes
but not always synonymous with primary audiences. The
latter fact means that the process of negotiating
research questions and other substantive issues may have
to involve a number of parties.

Primary Audiences for National Evaluations

Executive Policy Staff

This category includes individuals with authority ov r
resource allocations and the design of progr. s, most
importantly, senior-level agency officials and their
analytical staffs and budget examiners in the Office of
Management and Budget (OMB). It is rare, ir ever, that
these officials are waiting for evaluation study results
in order to make up their minds on what policies to
pursue or what programs to fund. The weight of an
evaluation may be slight in comparison to the
constellation of interests and other reasons for deciding
one way or another, even in ways counterindicated by an
evaluation study.
The temptations to misuse or not use the results of
evaluation studies are all too clear; hence the
importance we place in this and other chapters on the
obligation of evaluators to release findings
independently of executive decision makers. These
temptations are also the reason (as we indicated in

Chapter 3) for recommending that all evaluation studies
be subject to review, the results of which are made
public (see Recommendation D-6).

One of the problems in the utilization of evaluation
results is that findings may not be disseminated to all
persons involved in making decisions at the executive
level. This is often true for OMB staff, who are
generally not in the "loop" of people who normally
receive evaluation reports, so their information needs
may be served inadequately. In addition, turnover of
top-level agency officials in education has aggravated
the problem of loss of information and institutional
memory. On the other hand, agency officials have the
advantage of being able to draw on their policy and
evaluation staffs, who are probably the most consistent
users of evaluation data while also being the likely
immediate sponsors of evaluations.

The potentially short life of evaluation findings,
even though the knowledge might be useful at a later time
and in a different context, means that dissemination
should not be just a one-time effort. Archived
evaluation studies that are difficult to obtain and whose
existence is difficult to determine are useless. Hence
some attention should be given to the problem of
re-dissemination of evaluation findings; perhaps in the
form of summaries or reviews of past evaluation findings
for executive-level officials as programs and policies
come up for review.

Congressional Policy Makers

It is a mistake not to differentiate among congressional
users of information. Rarely are members of Congress
direct and immediate audiences. Rather, the initial
contacts are more often with the Congressional Research
Service (CRS) staff, committee staff, or personal staff
of members of Congress. In addition, staff of the
Congressional Budget Office and of GAO are frequently
prime audiences for evaluation studies. CRS, as part of
the Library of Congress, functions as a quick reference
service for both members and committees of Congress; GAO
carries out special studies at the behest of
Congress.[13] Congressional staff themselves differ in
their use of evaluation information: senior staff of
committees are generally better informed users of
evaluation results than personal staff of individual

members, who tend to be junior, must cover a much broader
range of issues, and must generally find evidence to
support a member's view. There are also differences
among types of committees: authorization committees tend
to cite evaluation data more frequently than
appropriations committees (see buruoh, Leviton, Cordray,
and Pion, Ch. 6:12-18o in Boruoh and Cordray 1980)--
proof, perhaps, of the fact that budgetary decisions
often are not heavily influenced by the results of
program evaluation.

It is relatively easy to document the explicit use of
evaluation studies by Congress and its staff: who makes
what information requests and receives responses from
CRS, who has received copies of evaluation studies, and
who refers explicitly to those studies in committee
reports and in the published remarks of members of
Congress. But there is also a more informal and diffuse
infiltration of information into congressional discourse
that is much more difficult to trace because it leaves no
explicit markers. Thus, a Congresswoman who remarks on
the floor that a particular program is working well may
mean that she has talked to a school principal in her
district who assured her that without the program his
schools would be suffering, or she may mean that she has
received a memo from one of her staff who had summarized
an evaluation report from the Department of Education, or
she may be referring to an assessment from GAO, or she
may merely be expressing her own opinion based upon
whether or not the program is "in line" with the kinds of
things she usually supports. We suspect, along with
others, that this informal, diffuse use of evaluation
results may be the most important use of all, but it is
not something for which one can readily provide direct
documentation.

Federal Program Managers

Program managers are likely to be interested in
information that can improve delivery of educational
services at the local levels. Since they are often
already committed to a given program, effectiveness
information may seem irrelevant to them except insofar as
it enhances support for the program. On the other hand,
information on how programs are being implemented and
what services are being provided to what beneficiaries
can lead to improvement in program regulation and

management. However, if the changes suggested by findings of process evaluations are too disruptive of established procedures, they are not likely to be implemented.

Recommendation D-12. The Department of Education should ensure that evaluations deal with topics that are relevant to the likely users. (See Recommendations C-1 and D-1.)

. As discussed earlier, relevance is not easy to achieve, but it is relatively easy to specify procedures that will make it more likely. Such procedures include:

* Primary audience(s) must be specified from the beginning of the study.
* Arrangements must be made to facilitate communication between evaluators and intended users at the inception of a study and throughout its course. This will help ensure the fidelity of the evaluation to the questions of interest to the identified audience(s) and will also help obtain commitment and interest on their part. Current administrative restrictions that inhibit that kind of communication should be removed.
* When the goal of an evaluation is to provide information for decisions at specified times, such as the reauthorization of programs or annual program appropriations, reports must be delivered on time. If a study has been delayed, its abortion should be considered unless some aspects will address longer-range concerns.
* Evaluation monitors should be charged with the responsibility of including in their routine monitoring information about events and changes that carry implications for the usability of findings. Changes in evaluation design or methodology are sometimes made in response to field conditions, budgetary and clearance constraints, or for other reasons. Such changes may have sufficient impact on a study so that the research questions framed to be relevant to the identified audience(s) can no longer be addressed adequately. Changes in the conduct of an evaluation that have such impact on the possibility of utilization should suggest rethinking the objectives of the evaluation or terminating it altogether.

133

Secondary Audiences for National Evaluations

Other audiences also have a stake in federal education programs, and therefore in evaluations of them, even if the questions addressed have been framed by the concerns of federal legislators or executives. Of course, some studies done at the national level may specifically address the information needs of a nonfederal audience, for example, representatives of minority and other beneficiary groups. For studies initiated by or at the behest of any of these other audiences, our classification of primary and secondary audiences would, of course, be reversed.

State and Local Agencies: Central Staff

The distinctions made at the federal level among decision makers, evaluation (and other analytical) staff, and program managers are also important at the state and local levels. The motivations and general information needs of the staffs are analogous, but focused on the program as it operates in the local setting. Since the policy variables that can be altered by state and local administrators are considerably different from those that can be altered by federal staff and Congress, evaluations must address different questions. Similarly, program management at the federal level entails quite different responsibilities from program management at the state and local levels, and process evaluations that are intended to improve management must be sensitive to these differences.

Local Agencies: Principals and Teachers

The individuals who actually provide the educational services intended by a program (and their representatives, such as the National Education Association (NEA), the American Federation of Teachers (AFT), and associations representing school principals) can become a powerful constituency for or against a program, as has been demonstrated by the history of Head Start and the experiments with voucher programs. Evaluations can be threatening or supportive--threatening if they appear to suggest a reduction in a program viewed as useful, supportive if they offer help to teachers in

doing a better job with a program. If the purpose of an
evaluation is to do the latter, then it must address
program elements that are under the control of teachers
or principals. For example, demonstrating differential
effects of a program for different population groups is
not helpful to teachers or principals since neither can
select whom they will teach. However, demonstrating
differential effects of alternative program strategies
may be helpful, since teachers can select the strategy
most appropriate to their school situation and students.

Program Clients and Their Representatives

The ultimate targets of education programs are students.
Since much of the investment in federal education
programs is at the elementary level, obviously many of
the beneficiaries are too young to be audiences for
evaluation information. However, there have been
specific attempts to address evaluations to parents so
that they could use the results to improve their
children's schooling. As we noted above, this was the
explicit intent of the original Title I evaluation
mandate (the first legislated requirement for evaluation
in education) as originally proposed by Senator Robert
Kennedy in 1965 (David 1978). The objective has seldom
been met, even when parent advice was legislated into
later Title I amendments in the form of parent advisory
councils. Groups other than parents also speak for the
interests of beneficiaries, most of whom are poor,
members of minority groups, handicapped, or otherwise the
targets of discrimination. The interests of these
groups, which include the major advocacy organizations
concerned with equal opportunity and minority issues, is
to use evaluation information to ensure that the intended
beneficiaries are adequately reached by the programs
intended to serve them and that those programs deliver
effective services.

Researchers

The outcomes of any evaluation study will be of interest
to other evaluators and researchers who are concerned
with development of educational policy, with
instructional strategies and school management, and with
the technical issues arising in the conduct of applied

research. Although clearly a subordinate audience,
evaluators and researchers should have easy access to
evaluation reports. In addition, primary data should
also be available to researchers so that secondary
analyses and cross-evaluation analyses can be carried
out. The importance of providing for secondary research
is demonstrated by the Cook et al. (1975) reanalysis of
the "Sesame Street" evaluation that showed that, although
the target population--poor children--had indeed made
gains in reading readiness, as documented by the original
evaluations, the gap between them and more affluent
children had actually grown because the latter made
greater learning gains. In order to provide for
secondary research, reports and primary data and
publication of evaluation-related material should be
archived in professional journals and as monographs (see
Recommendation D-7).

Media

Discussions of evaluations are more likely to find their
way into professional and trade journals if results turn
out to be controversial. If the program being evaluated
is itself of sufficient interest, the controversies are
likely to be picked up by the more popular media,
newspapers, television, and radio. Obviously, these are
secondary audiences for evaluation results, but the way
in which evaluators communicate with them may make a
crucial difference in the reporting and interpretation of
what a program is all about and what evaluation is all
about.

Reaching Audiences

Recommendation D-13. The Department of Education should
ensure that dissemination of evaluation results achieves
adequate coverage.

Evaluation utilization has been assigned a high
priority within the Department, but utilization cannot
happen unless people have a chance to consider relevant
information. Therefore, it is important to establish
clearly that attention to dissemination is not a pro
forma exercise. Indeed, the agency must, through its
actions, indicate as great a commitment to dissemination

concerns as to resarch design, measurement, and
analytical procedures. Staff who prepare RFPs and
monitor evaluations and external contractors or grantees
must both understand that attention to dissemination is
not just a "boilerplate" requirement, but that
dissemination plans will be subjected to the same
scrutiny and assessment as are evaluation designs and
methodology.

At the very least, evaluation results must be
communicated (delivered) to the primary audience(s).
This requirement would seem self-evident, but it often is
not met. Contract clauses routinely forbid dissemination
before formal approval by the sponsor, which is sometimes
withheld. As Boruch, Cordray, and Pion note (Ch. 5 in
Boruch and Cordray 1980), this keeps some (though not
all) evaluators from reporting on their findings. Also
routinely, a very limited number of copies of final
reports are printed (100 copies for most studies unless
unusual circumstances exist), with the result that
landmark studies like the Title VII bilingual education
study (Danoff 1978) quickly become out of print. In some
cases, a copy of the final report cannot even be found in
the project files (Cook and Gruder 1979). In other
cases, like that of the NIE compensatory education study
(National Institute of Education 1977), a stockpile of
copies actually exists, but it is difficult to get
information about how to get copies. In cases of lengthy
reports with multiple appendices, archives like ERIC
contain only part of the material originally published.
Restrictions on the number of copies and on archives--not
to mention more costly dissemination strategies--are
often imposed by contracting rather than technical agency
staff in order to reduce budgets but without
consideration of dissemination needs.

All RFPs and grant announcements should include
requirements for a dissemination plan that is oriented
toward maximizing the likelihood of utilization. The
evaluation of proposals should give appropriate weight to
the quality of the dissemination mechanisms proposed.
Budget negotiations should recognize that adequate
dissemination is costly and cannot be an afterthought.
Dissemination plans should include:

  • Specification of primary and secondary audiences;
  • Delineation of the different information needs of
the specified audiences and how those needs will be

served, such as different types of reports including more
or less technical material;
  • Provision for an adequate number of copies of
reports and other salient material to be distributed to
the specified audiences;
  • Strategies for reaching audiences through means
other than printed reports, e.g., conferences, throughout
the course of the study;
  • Specification of timetable events, e.g.,
congressional hearings, that provide occasion for
reporting on findings;
  • Mechanisms for reviewing and revising the
dissemination plan during the course of a study to take
account of changes in the study or in the context of the
work;
  • Plans for archiving reports and other
documentation of findings so that they remain accessible,
with a guarantee by the contractor that data will be
clean and accessible (see Recommendation D-7); and
  • A budget commensurate with the proposed
dissemination activities.

Recommendation D-14. The Department of Education should
observe the rights of any parties at interest and the
public in general to information generated about public
programs.

  Though minimal dissemination is concerned primarily
with the immediate or primary audience, other people
having an interest in the program being studied are
likely to demand and should have access to evaluation
findings. This raises two issues: What are the special
rights, if any, that should be afforded the agency that
has requested and funded an evaluation, e.g., the
Congress, the Department, OMB, or GAO? To what degree
should traditional authority relationships be overridden
in order to serve the public interest, i.e., what
obligations do evaluation units and contractors have to
disseminate findings to potential users who are outside
the command and report lines within tables of
organization?
  Findings from evaluations must be made available to
those who are importantly affected by the programs being
evaluated: for example, those who manage them, those who
provide program services, and those who are intended to
benefit (or their representatives). Since evaluations

are paid for with public funds, they should also be made
available to the public at large. The Committee is aware
of the dangers in providing too much autonomy to
evaluation units and contractors, but public interest
needs suggest that, at the dissemination stage,
evaluators should be guaranteed a certain degree of
autonomy.

Four steps are needed to provide improved public
access to evaluation findings:

• Proper safeguards for maintaining the rights to
privacy of individuals and organizations must be applied
before release of findings;
• The rights of the sponsoring authority to
exclusive access to evaluation results should be limited
in time;
• The right of managers and executives to restrict,
control, or suppress evaluation findings should be
limited in time; and
• Reports on findings should be accompanied, when
available, by interpretations and critiques issuing from
the review process recommended in Chapter 3.

Appropriate changes should be made in contract provisions
to allow contractors and grantees the necessary
flexibility with regard to distribution of reports and
other dissemination strategies.

Recommendation D-15. The Department of Education should
give attention to the identification of "right-to-know"
user audiences and develop strategies to meet their
information needs.

Perhaps the most neglected audience for evaluation
studies consists of program beneficiaries and their
representatives. We recognize that this neglect is not
so much intentional as it is produced by the very real
difficulties of defining this set of audiences in a
reasonable way. In order to more closely approximate the
ideal that all those having a recognized interest in a
program should have reasonable access to evaluation
results, the Department should consider dissemination of
evaluation reports freely to groups and organizations
that claim to represent major classes of beneficiaries of
education programs. Positive, active dissemination to
such right-to-know groups may include such specific

activities as ascertaining their information needs prior,
to evaluation design and during the evaluation, preparing
standard lists of groups and organizations to whom
evaluation results are disseminated routinely, and
seeking out comments and critiques of evaluation reports.

Since it is to be expected that such right-to-know
groups will be different for different evaluations,
careful consideration of the appropriate right-to-know
groups should be part of the dissemination plans that
contractors are asked to prepare as part of their
response to RFPs and grant announcements.

We recognize that this recommendation makes the whole
process of sponsoring and carrying out evaluations more
complex, but we consider the involvement of right-to-know
groups critical. They often perceive themselves as
having limited access to or insignificant involvement in
evaluation efforts that may be used for policy and
resource allocation decisions that concern them.
Furthermore, such groups can have an important influence
on the improvement of educational practice, and they need
access to information so that their recommendations and
actions are as effective as possible. Involvement of
these audiences from the very outset of an evaluation
enriches the public policy process both because it widens
the universe of viewpoints and because, over the long
term, it can improve the quality of education insofar as
these groups are links to the communities that the
government is attempting to serve. If they share in the
evaluation process from the beginning, they are more
likely to use the findings in their spheres of influence.

Changing User Behavior

Recently Sechrest (1980) has suggested that, if
high-level administrators could be trained in how
evaluations are done and how researchers present results,
utilization would be increased. We include suggestions
for such training in Recommendation D-17 in Chapter 5.
We have some doubt, however, that top executives or
members of Congress have the time for such training or
would retain technical knowledge that they would use
infrequently. If they did develop greater facility for
the language of evaluation, they would certainly become
more sophisticated readers.

It is possible to think of incentives for use and
sanctions against failure to use evaluation results

140

within the lower echelons of federal and local program
management. For example, program managers and program
personnel might be required to respond to evaluations
with appraisals and critiques, to provide plans for
incorporating valid findings into their program
operations, and to document subsequently whether the
planned changes had been made. Some states (Rhode
Island, Massachusetts) do indeed require reports from
local school systems on the use of Title I evaluations.
However, there is also some danger that such requirements
will turn into additional pro forma exercises. Required
responses and actions might also make explicit some
conflicts between managers and analytical staff about the
value of a program or the effectiveness of its management.

Recent reforms in the federal civil service provide
special bonuses for effective program management, and
appraisal of management is tied to the results of program
evaluation (Office of Management and Budget 1979).
However, the success or failure of a program is at least
as much dependent on its design and legislative
provisions as it is on the efforts of program managers
and personnel, so the attempt to judge good management
performance through program evaluation may be off target
unless only those factors under control of the program
manager are examined. A second effect of this particular
incentive system has been to define management objectives
in clearly measurable terms (e.g., items of priority mail
answered on time) rather than in terms of the more subtle
and less objectively measurable behaviors that are needed
for effective program management, such as frequent and
productive interaction with state and local staff.

Sanctions for failure to institute changes suggested
by evaluation results have also been suggested, for
example, withholding program funds until the changes are
made. The history of cutoff of federal funds for
violation of civil rights laws suggests that this
particular sanction is very unlikely to be imposed.
Consequently, we make no explicit recommendation on the
use of incentives or sanctions. However, the Department
might consider requesting that federal program managers
who have had their programs evaluated prepare evaluation
use reports. These might be prepared within one year
following receipt of the evaluation report and contain an
assessment of the level and types of uses made (including
reasons for nonuse) as well as an analysis of factors
that impeded or facilitated use. If the Department
proceeds with such a requirement, the dissemination and

linkage unit proposed above should be charged with the
additional responsibility of assessing whether drawing
the attention of program managers to evaluation
information in this manner actually improves its chances
for use.

NOTES

1 The literature on putting knowledge to use has grown
   as rapidly as the evaluation field itself. Davis (in
   Human Interaction Research Institute 1976) has
   estimated that, by the mid-1970s, the research
   literature concerned with the field of knowledge
   utilization included some 20,000 citations, compared
   with 400 such citations 20 years earlier.
2 For example, Marsh et al. (in press) found that
   changes in rape law had produced a statistically
   significant decrease from 12 to 10 in the average
   number of examination procedures that a rape victim
   had to undergo if she reported the crime. Obviously,
   in substantive terms of victim humiliation, one could
   hardly report this as a meaningful change.
3 The average tenure of a Commissioner of Education
   during the last decade has been less than 2 years; NIE
   has had six changes of leadership in 8 years.
4 We analogize from a definition by Yin et al. (1976) of
   situations regarding the adoption of innovations:
   adoption is regarded as a positive outcome if the
   innovation leads to improvement but as a negative
   outcome if it does not; failure to adopt is a negative
   outcome only if the innovation would indeed lead to
   improvement but a positive outcome if it would not.
5 Head Start teachers deciding to increase the time
   spent on prereading activities are as much decision
   makers in their realm as a superintendent installing a
   new curriculum, a state legislature passing an
   appropriation for compensatory education, or a federal
   program manager developing program regulations.
6 Of course there is always a question as to who can
   represent beneficiaries. The Committee has made no
   attempt to address this question in any detail, both
   for lack of time and because we did not consider
   ourselves qualified to define such representatives.
   We note that there are groups that speak on behalf of
   specific beneficiary groups; their claims to represent
   these groups could, perhaps, be considered in the same

light as the claim of public officials that they
represent the public.

7 Evaluations done by individuals or units that also
have operational responsibility for a program are
generally mistrusted. How much more objective
evaluation becomes when it is done by third parties,
but still under the auspices of the program, is not
clear, particularly when future evaluation contracts
from the same source are a possibility. Evaluations
performed or sponsored by units outside a program are
not necessarily free of bias either, whether performed
in-house or contracted out, especially when top
decision makers are known to favor particular points
of view.

8 Appropriate packaging has also been deemed important,
but many counterexamples exist. For example, the
attempt to develop social indicators resulted in a
handsome publication (Office of Management and Budget
1973, U.S. Department of Commerce 1977) with
attractive and easy-to-read graphics, yet it has found
limited use.

9 As we discussed above, there are risks for
bureaucracies of having to deal with new information.
Other groups also run risks: for example, audiences
concerned with equal educational opportunity may find
negative results on programs they favor distasteful
and disturbing.

10 The distinction is not always clear. Sometimes,
expectations for use at all levels are set up when
data required at the federal level are collected by
staff at the local level, as in the case of Title I.
In some cases, it may be most efficient to sponsor a
study at the federal level even when the results are
pertinent to individuals at the local level; for
example, testing the efficacy of alternative
strategies for teaching reading.

11 The national-level evaluation of ESEA is not intended
to take the place of the three-tier evaluation of
Title I based on local data collection and aggregation
at the state and national levels. Rather, it is a
substitute for previous efforts at the national level
to study the effects of Title I, specifically, the
sustaining effects study (Dearman and Plisko 1979,
U.S. Department of Health, Education, and Welfare
1979a, Baker and Ginsburg 1980).

12 As described in Appendix A, fiscal 1980 was the first
year for which there was a comprehensive review of

143

evaluation plans from different components of OE, and
that review did not include the relevant activities of
NIE and the National Center for Education Statistics
(NCES). The new Department has attempted to institute
a more centralized evaluation planning system; at this
time, one cannot gauge the degree of its
implementation or success.

13 The changing role of GAO and its success in responding
to new demands have been described by Levitan and
Wurzburg (1979) and by Mosher (1979). Though Congress
broadened GAO's mandate as early as 1945 to include
monitoring of the administration of programs as well
as of expenditures, it was not until 1967 that GAO
became active in the field of program evaluation: a
review of OEO's antipoverty programs was its first
effort. In the succeeding decade, GAO has been
changing its staff and organizational structures in
order to carry out with greater effectiveness the
increasing number of program evaluations undertaken by
the agency. At present, studies carried out by GAO
range from investigations of misallocation of funds
within government agencies to impact evaluations of
social programs and even to the evaluation of
evaluations carried out by executive agencies (U.S.
General Accounting Office 1977, 1978).

5
Organizing and Managing
Evaluation Activities

Many of the issues of quality and utilization discussed
in the preceding two chapters are related to the way in
which federal, state, and local education agencies
support and sponsor federally funded evaluations.
Dealing with those issues requires consideration of three
major factors:

* Responsibility. What kinds of evaluation
activities is the Department expected to carry out as
part of its oversight functions and of its effective
management of federally funded education programs? What
should it do for effective policy formulation? What
ought to be the responsibilities of local and state
education authorities?
* Organization. How are the evaluation activities
now organized in the Department and why? How should
those activities be organized in order to maximize
capabilities and incentives for producing reliable
information and high-quality analysis?
* Constraints. What are the impediments to
producing evaluations of high quality and using results
effectively? Which of the impediments are the result of
external constraints and which are due to internal
procedures? Which of the external constraints can be
alleviated? How can internal processes be improved?

Discussion of these issues reinforces a number of the
recommendations made in earlier chapters. In this
chapter we suggest guidelines for balancing the need to
decentralize and to coordinate evaluation activities; we

133

also make some additional recommendations on improving
the management of evaluations.

RESPONSIBILITY FOR EVALUATION ACTIVITIES

In Chapter 2 we discussed in general terms the different
types of policy questions that are asked about
established or proposed programs. In this section we
consider what kinds of evaluations need to be carried out
in order to address those policy questions for education
programs.

Accountability

The Department is accountable for carrying out education
laws in three respects: ·ensuring that moneys are
allocated as specified, ensuring that benefits go to the
targeted groups, and ensuring that civil rights
provisions and service mandates are being met.

Fiscal Accountability

Because of the decentralization of education, the
allocation of funds for most major programs takes place
at all three levels of government: federal, state, and
local. (A few programs provide for federal grants
directly to local agencies.) Hence, all three levels
must account for the use of federal education funds, and
fiscal reports from local and state agencies form the
basis for the Department's own fiscal reports. Grantee
reporting is checked periodically by the agency's
inspector general. For a few titles, like vocational
education state grants, such auditing is mandatory in
law; for the most part, however, the Department has
discretion as to what local and state reports and
disbursements are audited. Nearly one-fourth ($10
million) of all evaluation funds are spent on fiscal
audits; generally, programs with large outlays (Title I
of ESEA, post-secondary grant and loan programs) receive
most attention (see Appendix A).
   As audits have gone beyond checking for sound fiscal
management and into checking for compliance with legal
requirements on the use of funds, the line between fiscal
audits and other accountability evaluations has become

146

fuzzy. For example, whether Title I money is used to
supplant regular school funds or provide supplemental
services to eligible participants (Martin and McClure
1969, Stanford Research International 1977a, 1977b) has
become an issue affecting the substance of what goes on
in the classroom. The early problems with supplanting
have caused most school systems to provide "pullout"
programs that can be easily accounted for separately,
even though they may not be the preferred educational
option (National Institute of Education 1977).

Accountability for Beneficiary Coverage

Grantee reports have generally served as the most
comprehensive source of information on program
participation. Though local agencies are obviously in
the best position to count participants, there are two
problems with the use of such self-reporting:
reliability of the reported data and lack of information
on who is not being served. Reliability can be
documented through third-party checks on grantee
reports. If grantee reporting for a specific title turns
out to be highly unreliable, technical assistance on
interpretation of the law (e.g., defining participants
properly) may be warranted; alternatively, incentives and
sanctions that encourage misinterpretation need to be
examined and adjusted to bring grantee performance and
reporting in line with the legal intent. It is doubt
that the Department will ever be able or wish to rep
grantee reporting on beneficiary coverage, but it must
accept responsibility for the accuracy of such reporting.
    How to document the number of potential beneficiaries
not being served is quite another matter, however.
Establishing the universe of eligible participants falls
under the heading of needs assessment. The incentives
and disincentives for conducting accurate needs
assessment may be strong at the local and state levels:
there is an incentive when having more eligible
participants means getting more federal dollars; there is
a disincentive when federal dollars are accompanied by
matching provisions that call for greater contribution
from local and state than from federal sources. At the
federal level, there are also strong incentives: program
administrators who do not want to see their programs grow
are rare indeed,[1] yet this responsibility is often
assigned to a program office, as was the case in

developing P.L. 94-142 (Education for All Handicapped
Children Act). Because of the incentives, we conclude
that needs assessment ought to be carried out not by
program offices, but by parties with no stake other than
accuracy in the outcome. The cooperation of local and
federal program managers is necessary, however, since
needs assessment must be informed by intimate knowledge
of the local context and of potential program benefits.

Accountability for Civil Rights Mandates

Accountability for civil rights mandates takes two
different forms in education. The first involves the
enforcement of civil rights statutes in any way related
to educational institutions, whether built into federal
education legislation or decreed by federal courts, and
is based on federal responsibilities under the
Constitution. At the same time, the provision of
educational services is constitutionally a state
responsibility, delegated to local authorities.
Enforcement of statutes relating to civil rights and
equal educational opportunity has become the
responsibility of the Department because it can withhold
federal funds in the event of noncompliance. As with
fiscal accounting, a separate office headed by the
Assistant Secretary for Civil Rights is responsible for
compliance, and it is not considered an evaluation
activity per se.

   The second form of accountability arises because some
civil rights statutes require certain kinds of
educational services. Two groups are specifically
covered in this manner: all handicapped children are
entitled to a free appropriate public education under
P.L. 94-142, and Title VII of ESEA (in accord with the
Lau court decision) requires schools to provide
instruction that does not put a non-English speaking
child at a disadvantage. Such educational services that
are spelled out in laws or in regulations tend to be
based on perceptions of constitutional rights rather than
on social science evidence about needed services.
Consequently, monitoring activities may overlap.
Responsibility for compliance with service mandates may
belong to the program office, but selective checks are
often carried out by the Office of Civil Rights. An
example is the labeling and testing of handicapped
children.[2] Since these two kinds of offices tend to

respond to different constituency groups (minorities and other targets of discrimination on one hand, school systems and educational institutions on the other), they generally have distinctly different views of what ought to be expected of grantees. Overlap of responsibilities is not undesirable if it is included in overall evaluation planning; otherwise it leads to inefficient use of resources at best and antagonism between units of the Department at worst.

Program Implementation

Except for provisions connected with civil rights and equal educational opportunity, federal education legislation often does not spell out mandatory educational interventions or treatments. The constitutional delegation of responsibility makes decisions in education a jealously guarded right of local and state authorities. Exceptions are such demonstration programs as Follow Through or Experience-Based Career Education, in which school systems are given the choice of one of several specified curricula. Since the rationale of demonstration programs is developing and testing effective interventions, documenting the nature of the services provided through them ought to be an integral part of any evaluation research associated with them. There are also some ESEA titles that include explicit process specifications, such as the requirement for developing an individual education plan (IEP) for every handicapped child served under P.L. 94-142. In the case of such mandated educational processes, especially those instituted on little evidence as to their effects, more than mere compliance checking is also needed. Evaluation should be carried out to find out the degree to which such processes contribute to the overall, goals of the legislation, for example, to provide more effective education for handicapped children or—in the case of bilingual education—for children whose native language is not English. Documentation of program process and implementation has been carried out at all government levels and, within the Department, by both the cognizant program units and the central evaluation unit.

V.

Program Effects

·With few exceptions, federal funds allocated at the
elementary and secondary levels are intended to bring
about improved education for those students who have not
been served·adequately in the past. Because the total
amounts spent are large,[3] Congress from time to time
has called for information on program effects. In the
past, the response by OE has been the commissioning by
the central unit of large-scale impact assessments that
consume several years and millions of dollars, as
exemplified by the sustaining effects study carried out
by the Systems Development Corporation (1976, Baker and
Ginsburg 1980). There have been several problems with
such efforts. First, what Congress often wants and needs
is information on effective delivery, in the sense of
having accurate accounting for how a law is being carried
out, as described above. Better specification of the
questions to be answered in any legislation calling for
assessment (as recommended in Chapter 2) would help avoid
misdirected evaluation efforts; even more important is an
ongoing dialogue on congressional needs between key
congressional staff and Department staff responsible for
evaluation.

Second, even when assessment of program effects is
called for, expectations of the size of those effects are
often exaggerated because of unrealistic promises during
the legislative and appropriation processes. But by the
very nature of federal education programs, effect
expectations should be modest. Whatever educational
service is envisaged as a result of federal dollars, it
will be delivered in a decentralized manner through some
16,000 local school systems in the public sector
comprising nearly 90,000 school buildings. There are
more than 2 million teachers in the public school
systems, and another 250,000 people are teaching the 10·
percent of students in nonpublic schools. (Private
school students also receive benefits under Title I and
other federal programs.) Federal programs operate at the
margins of this huge enterprise, providing 8 percent of
all revenue for public elementary and secondary schools
(Dearman and Plisko 1979). Moreover, most federal
programs are geared to specific populations; in those
cases, support for core education, the major
responsibility of the teacher, is expressly ruled out.
Yet the children who receive benefits from any of the
federal titles do not do so in isolation from the rest of

their education. Finally, as we have noted, federal
programs more often than not have multiple and
amorphously defined outcome goals, though they are
usually explicit regarding distribution of benefits. To
expect strong treatment effects under these
circumstances--for example, improvement throughout the
country in school achievement of a target group or
lessening of racial tensions--is to ignore the nature of
the educational system in this country.

When the effects of a given program are modest, their
estimation is a complex, difficult, and costly task.
Such estimation should be done only when it is likely to
affect program decisions (for example, in the case of a
limited experimental program) and only by the most
competent evaluators and evaluation organizations.

Program Planning and Improvement

One of the Department's responsibilities is to provide
leadership for improving education in this country;
therefore, it ought to carry on a set of prospective
activities designed to improve the substance of existing
programs and to develop new programs. The relevant
evaluation activities are summarized in Chapter 2: needs
assessment, identification of interventions likely to
relieve the need, small-scale testing of proposed
programs under optimal conditions, field evaluation under
actual operating conditions, and analysis of likely costs.

Such a process of program planning should operate both
at the national level and in selected states and
localities that have the resources. A similar set of
activities is relevant to program improvement, although
the need and the general nature of the program may
already be established. Too often, however, the
exigencies of the budget process and the demands from
those concerned with implementation of current programs
relegate the planning of new programs and the improvement
of established ones to a low priority. The tracing of
benefits already legislated and the assurance that
programs are carried out as intended take first
priority. Development of knowledge needed to formulate
better programs is a long-term process, with no assurance
that the outcomes will be immedately applicable. In view
of pressures for greater accountability and improved
program management, it may be argued that activities
aimed at the substance of programs should be relegated to

the research component of the Department, but such an
assignment may lead to unfocused research not easily
related to program variables that can become part of a
federal education program. An interesting example of
coordinated program improvement research exists for
bilingual education, for which NIE, the program office,
and the central evaluation unit all participate in
evaluation and research planning. This kind of
coordination recognizes that, particularly for existing
programs, program managers should be involved in the
design and testing of alternatives. They can provide the
necessary experience regarding current program
operations, and they are likely to have ideas for
improvement. But the overall effort should be in the
hands of research-trained people whose full-time
attention can be devoted to evaluation activities.

Evaluation as a Management Tool

In an examination of the use of social science
information by federal executives, Caplan (1976) found
that, in the Office of Education, more program evaluation
was conducted and less of the information generated was
actually used than in any other agency examined. It may
be that, in its past emphasis on rigorous studies of
program effectiveness, the central evaluation unit of the
Department was not satisfying the information needs of
the most powerful audiences, namely, the legislative and
executive branch overseers. Their primary interest is in
fiscal and beneficiary information, which provides an
effective tool for holding managers at all
levels--federal, state, and local--accountable for proper
distribution of benefits. In fact, OMB circular A-117
(Office of Management and Budget 1979) requires both
management and program evaluation of every agency
(including an annual report) and ties this activity
directly to the reward system for federal managers
included in the recent civil service reforms.

   Problems are likely to arise, however, when
accountability demands are taken beyond ensuring that
resources are properly allocated. Who is to be held
accountable for program effects that will probably be
modest and difficult to estimate? As Cronbach et al.
(1980) point out, condemnations of individuals for
weaknesses or "failures" that occur in a system over
which they have little control is a perversion of the use

of accountability. The delivery of federal education
programs is a case in point. Given that authority is
dispersed and delivery of educational services highly
decentralized, it is difficult to assign responsibility
for program outcomes to specific institutions, let alone
to sets of individuals such as teachers, superintendents,
or federal program managers. This is not to argue that
studies of program implementation and of program effects
should not be done, only that they are unlikely to be a
useful management tool.

There is a second problem with using evaluations of
program effects for trying to improve program
management. The fear that programs will be curtailed
because of negative findings is aggravated in today's
climate of tightening budgets. Even if in the past there
have been few examples of established education programs
that have been cut severely or abolished as a result of
evaluation findings, the threat is real. Line managers
and top officials wanting to build programs and budgets
are not likely to cooperate enthusiastically in
evaluations they perceive to have the potential of
damaging their programs.

CURRENT ORGANIZATION

How effectively is the Department now organized to carry
out its evaluation responsibilities? Figure 3
illustrates the organization of the Department as of
January 1981; Figure 4 places the central evaluation
unit, which carries major but not sole responsibility for
evaluation, in its current context.

For evaluation activities other than fiscal accounting
and civil rights enforcement, legislation and
administrative actions have created a hodgepodge of
evaluation responsibilities and assignments, based more
on the power base and history of individual programs than
on rational planning. After an analysis of major
education programs, Cordray, Boruch, and Pion found:
"Programs differ markedly with respect to the number and
types of evaluation mechanisms that are described within
the law and by federal regulations" (in Boruch and
Cordray 1980, Ch. 3:7). Thus, states and localities may
or may not be charged with producing performance reports,
doing needs assessments, and carrying out studies of
program improvement and program effects. For some
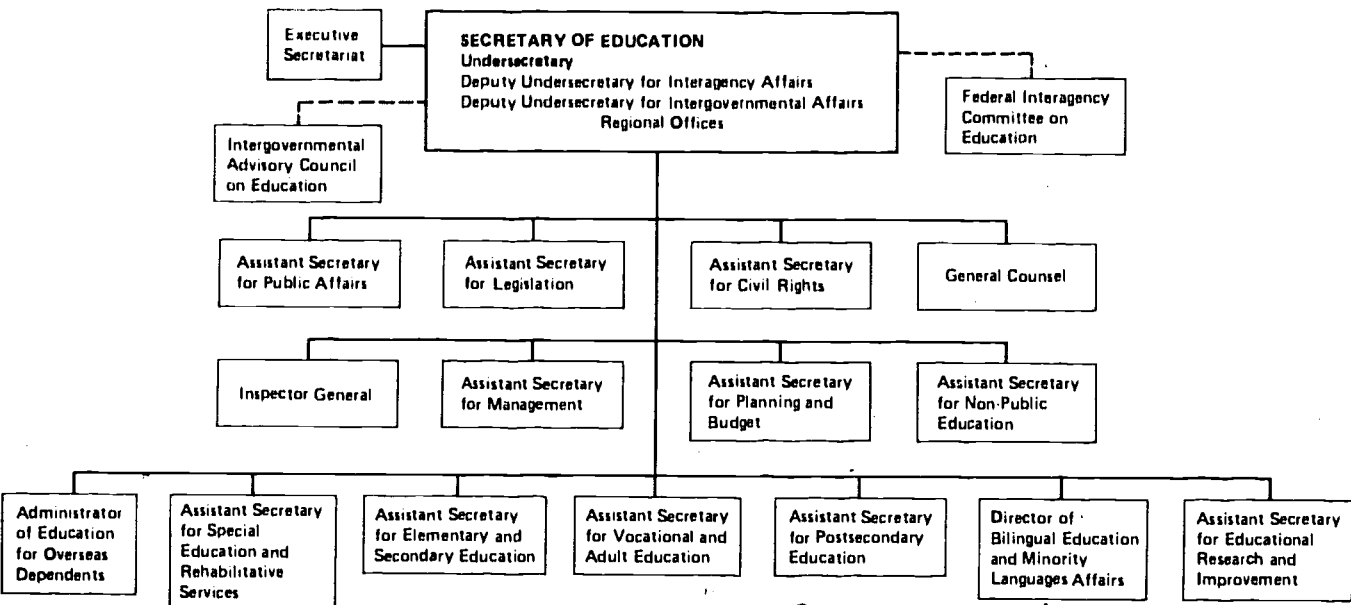programs, states are supposed to monitor local programs

FIGURE 3   Organization chart of the Department of Education.

154
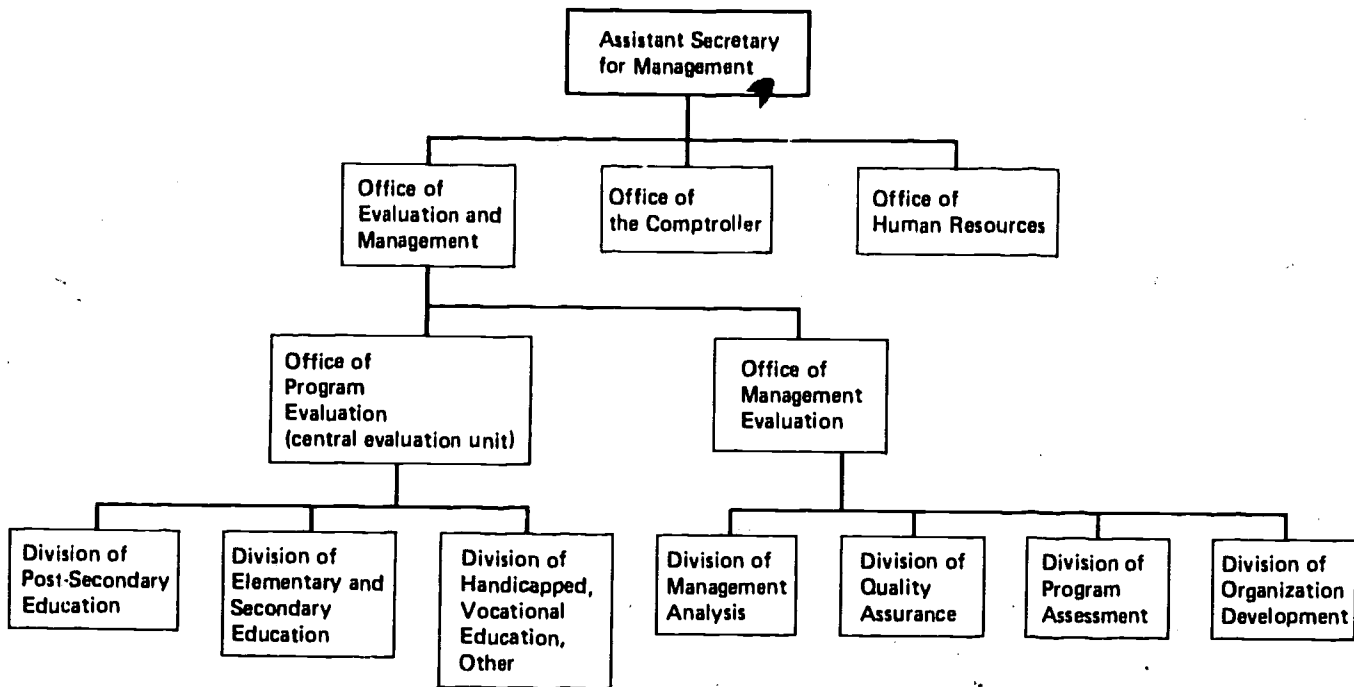
FIGURE 4   Organization chart of the Department of Education (Assistant Secretary for Management).

or local evaluation plans or both; for others, there is
no provision for review of local evaluations. Both
Congress and the Department have been responsible for the
present mix: Congress has attached dissimilar evaluation
requirements to various categorical titles that
distribute evaluation responsibilities differently from
program to program; the Department (and its predecessor)
have distributed evaluation responsibilities as much on
the basis of the political strength of individual program
administrators and their constituencies as on any basis
connected with the quality or integrity of evaluations.

There has been a central evaluation unit at the
national level for a decade, but its responsibilities
have varied, even as funding has increased (see Appendix
A). After the unit was established in 1970, evaluation
activities began to be centralized. The central unit
acquired staff, a budget, and responsibility for national
studies. This centralization was instrumental in
introducing rigor, integrity, and visibility to the
evaluation efforts mandated by Congress and sponsored by
OE. For several years, budgets and responsibilities
increased. But as dissatisfaction developed with the
perceived lack of timeliness and relevance of some of the
studies--not to mention unhappiness with some findings
deemed potentially damaging--pressure increased for
certain programs to be responsible for their own
evaluation activities. At present, some programs include
virtually no evaluation activities other than obligatory
program monitoring; others delegate evaluations to the
central unit; still others conduct all their own
evaluation activities. In addition to the central unit
and program units, evaluation activities are also carried
on in the research unit (Assistant Secretary for Research
and Improvement), the planning unit (Assistant Secretary
for Planning and Budget), and at the Secretary's level.
Until 1979, there was no overall evaluation planning or
coordination of evaluation.

Congressional restiveness with the performance of this
nonsystem led to still another layer, mandated
congressional studies to be carried out by a designated
unit: NIE in the case of studies on compensatory
education and on vocational education, NCES for a study
on discipline in the schools (P.L. 93-580), and the
Secretary's office in the case of a study on school
finance.

156

GUIDELINES FOR ORGANIZATION

It is neither necessary nor even desirable that the
organization of evaluation activities be precisely the
same for each education program. But the current
accretion of idiosyncratic evaluation legislation and
internal assignments originally made for political and
administrative reasons bears reexamination in the light
of some reasonable criteria, such as: the type of policy
question to be asked and the information needed; the most
effective and efficient ways of obtaining the needed
information; the intended use of the information
(likelihood that use will occur may depend on how and by
whom the information is generated); the size and nature
of the program; and the research capacity of the unit
considered for assignment of evaluation responsibility.
The application of such criteria will indicate what
changes might be made to improve the current organization
of federally funded evaluation activities related to
education. But since there is no one best way to
organize these activities, the implications the Committee
has drawn from the preceding discussion are presented
below as suggested alternatives rather than as
recommendations.

Centralization Versus Decentralization

Organizational researchers and management experts have
debated the merits of centralized organization compared
with those of incrementalism and mutual adjustment
brought about through coordinative mechanisms among many
autonomous units. Each form of organization has its
costs as well as its benefits. Central organization can
lead to more coherent activity, but it is time-consuming
as the decision process works up through the hierarchy
and back down for execution. It may also seem capricious
and arbitrary, especially in complex situations and
situations of uncertainty. Such conditions are
characteristic of most evaluation planning related to
social programs. On the other hand, while decentralized
planning and execution can come closer to satisfying
needs of individual units at the federal, state, or local
level, it can lead to duplication, wasteful use of scarce
human and fiscal resources, and low quality. Attempts to
minimize these negative consequences through purposeful
coordination will, like other centralizing mechanisms,
exact high costs in time.

157

The Committee believes that the different evaluation
questions that need to be addressed concerning federal
education programs are now so diverse and of suoh varying
importance to different audiences that decentralization
is warranted. But responsibilities should be assigned in
a somewhat more planned manner than at present. There is
agreement within the current Office for Management, which·
has overall responsibility for program evaluation, that
some evaluation activities need to be decentralized; in
fact, present law and custom so dictate. But planning
directives for 1980 manifested an attempt to reoentralize
evaluation activities through review and approval by the
central unit of all evaluation plans. No parallel
attempt is evident with respect to evaluation activities
funded by federal funds at the state and local levels,
except to provide technical assistance in the case of
Title I evaluations.

Decentralization Among Levels of Government

As noted, evaluation requirements levied upon local and
state agencies vary from program title to program title.
(For summary descriptions of requirements in major
titles, see Cordray, Boruch, and Pion, Ch. 3 in Boruch
and Cordray 1980). Generally, reporting requirements
appropriately emphasize the collection of information on
beneficiaries served and on distribution of resources.
For a number of titles, the states carry the
responsibility of aggregating data provided by each local
education agency. But state-level reports have seldom
been able to make statements about how programs operate
throughout the state as a whole, partly because local
agencies were not reporting data of sufficient quality
and uniformity to allow aggregation. Consequently,
states have also acquired some responsibility for
technical assistance. For certain titles, localities are
also required to identify the number of individuals in
the target population (for example, for the handicapped
covered in P.L. 94-142). Since identification of
individuals generally leads to the need to serve them,
and federal funds by no means pay the total cost of
service, there are considerable disincentives to
comprehensive needs assessment carried out by local
agencies.
    In addition to reporting on the distribution of funds
and on the numbers and types of both potential

participants and those actually served, some titles
require reports on "effectiveness." In most cases,
effectiveness turns out to be the degree to which the law
is being implemented, i.e., whether program services are
being provided as specified in law and regulations. A
few local and state agencies also carry out evaluations
concerned with educational effectiveness. In many cases,
however, major expenditures of their own funds reported
by local agencies as evaluation of program effectiveness
are for testing designed to track general student
achievement rather than specific effects traceable to any
one program. It appears to be the intent of current
requirements that local evaluations serve auditing and
monitoring purposes while at the same time also informing
local program developers and administrators on the best
implementation strategies. As illustrated by the history
of Title I evaluations (summarized in Chapter 4),
stipulations for local and state evaluation activities
have shown a confusion of purpose between assessing the
extent to which programs are providing benefits and
mandated services and determining ways in which local
programs might be improved. Local evaluators are forced
to use designs and methods to collect data that can be
aggregated at the state and national levels, but such
data do not serve the local needs well. Moreover, those
data have not even proved useful in providing statewide
or nationwide overviews; separate state or national
studies have been needed for that purpose. Though some
data collected at the local level might serve both local
and national purposes, each type of evaluation question
has distinctive design and measurement requirements (as
discussed in Chapter 2) and implies different
relationships among the three levels of government.

We have noted in Chapter 3 the variable quality of
evaluation activities carried out at the local and state
levels and have recommended that Congress consider a
diversified strategy of evaluation requirements at these
levels (Recommendation C-3). In Chapter 4 we discussed
the need to build in the concerns of target audiences
from the beginning to increase the likelihood that
evaluation findings will be used. Consideration of how
scarce evaluation resources can be best employed to yield
reliable information that is useful to the maximum number
of audiences reinforces the notion that division of
evaluation responsibilities deserves more careful
analysis than it has received.

All grantees receiving federal funds for education

programs have stewardship responsibilities. At a
minimum, therefore, all such grantees should continue to
be required to report on the allocation of funds, on the
numbers of beneficiaries served, and on compliance with
the law where services and processes are spelled out.
But considerably more thought should be given to the
amount of such information that can be digested at the
state and the federal levels. The impression persists
that grantee application and reporting requirements are
intended to cover all bases and collect every conceivable
bit of information, creating such an overload that most
of the data pour in without being scanned, let alone
used. For example, in the migrant education program, OE
required the states to send copies of all subgrants to
OE. According to the program auditors, this mountain of
information simply collected dust in a storage area with
no attempt made to review it (Rock 1980). The practice
was ended as a result of the program audit. More
carefully considered requirements would reduce costs and
response burden and provide fewer and briefer reports
more likely to be reviewed.

Requirements that go beyond the basic reporting needed
for accountability functions should not be levied on all
localities and states alike. Questions on how a program
actually operates in the school, questions on the
detailed nature of the services and variations in
different localities, and--most difficult of
all--questions on the educational effects traceable to a
specific program need not be answered by all localities
or grantees. Cost effectiveness questions dealing with
the desirability of different program alternatives are
probably an even less appropriate requirement at the
local and state levels. Scarce evaluation resources are
frittered away when demands are made of all that could be
responded to more effectively by selective sampling in
nationwide studies or by studies carried out by
individual local systems or states with proven competency
and sufficient fiscal and human resources to evaluate
their own programs. These considerations lend additional
force to the recommendation made earlier:

Recommendation C-3. Congress should institute a
diversified strategy of evaluation at the state and local
levels that would levy minimum monitoring and compliance
requirements on all agencies receiving federal funds, but
allow only the most competent to carry out complex
evaluation tasks.

To this, we add a recommendation regarding the
Department's responsibility.

Recommendation D-16. The Department of Education should
clearly spell out minimum requirements for monitoring and
compliance reporting and set standards for meeting the
requirements.

The objective of this recommendation is to improve the
quality of data needed for accountability without
increasing the burden of response on local and state
agencies. Such data items as distribution of funds,
number and types of beneficiaries being served, and
specific program services should be defined by the
Department so that local and state agencies know exactly
what reporting is required of them. Quality control
procedures should be enforced so that performance reports
can be made to Congress. Before setting the
requirements, however, the Department needs to examine
its own capacity to deal with local and state reports so
as to avoid collecting information that is never used
because of the sheer inability of federal staff to deal
with the volume.
   In order to assist agencies in complying with federal
reporting requirements, the Department should extend
technical assistance as recommended above (Recommendation
D-8).* One way to provide such assistance would be to
select local and state agencies doing an exemplary job of
reporting. If none exists, the Department should fund
the development of such examples. Care must be taken to
select different types of locales exhibiting a variety of
student, teacher, and resource mixes. The exemplary
procedures should then be actively disseminated through
existing channels, for example, the Department's regional
offices, the Title I TACs, the NDN, or the state agencies.
   A second way to provide technical assistance would be
to make funds available to selected exemplary local
agencies to provide technical assistance on meeting
reporting requirements to less skilled school systems of
comparable type--something like the
"developer/demonstrators" funded by the NDN (Far West
Laboratory for Educational Research and Development 1979)
to provide training, materials, and technical assistance
for adopting exemplary education programs. After the
first 2 or 3 years, such funding should be based on the
success of an agency designated to provide technical
assistance in improving the reporting of those receiving
the assistance.

Decentralization Within the Department

Different evaluation activities are appropriately located
in different units of the Department to take advantage of
incentives for using results and of staff interests and
competencies. Using the typology developed in Chapter 2,
we suggest general guidelines for locating evaluation
activities within the Department.

The Office of the Inspector General should continue to
monitor whether funds are distributed according to law
and are allocated for the prescribed purposes. When
questions arise as to whether such additional services as
the law mandates are being provided to the target
population(s) (rather than the funds being used for
regular school operations), they need to be investigated
through evaluation strategies and methods appropriate to
documenting the nature of program interventions. This
type of evaluation requires research capabilities beyond
the scope of the Office of the Inspector General.

Accountability questions on beneficiaries served and
on program delivery should be monitored by officials who
administer the programs at the federal level, namely the
Assistant Secretaries for Elementary and Secondary
Education, for Special Education and Rehabilitation
Services, for Post-Secondary Education, and for
Vocational and Adult Education and the Director of
Bilingual Education and Minority Languages Affairs.
Responsibilities should include the monitoring of program
coverage and of provision of services mandated by law and
regulation (including such associated requirements as the
setting up of parent advisory councils). Where civil
rights laws a involved, the Assistant Secretary for
Civil Rights ha and should continue to have
responsibility. Much of the information on program
coverage and delivery should be obtainable through
focused grantee reporting using adequate quality control
and technical assistance measures, as discussed above.

There is continuing need for a central evaluation unit
to carry out activities not directly linked to program
accountability. First, the unit should sponsor, on a
sample basis and in cooperation with the program unit,
documentation of program process and detailed
implementation so as to provide insight on how
educational services have been changed. Second, also in
cooperation with the cognizant program units, the central
unit should support program improvement or development
studies, including needs assessment and understanding of

program context, the testing of promising alternative
program strategies, and analyses of the effects of
proposed changes in law or regulation. Third, when the
issue is educational effectiveness, the unit should carry
out--in cooperation with the program offices--needed
evaluability studies to define objectives and appropriate
measures. Only if such measures can be successfully
established and only if a program is of the type and at a
stage to allow impact evaluation (see Chapter 2), should
such a study be undertaken and then only if the need for
it can be justified.

The reason for assigning shared responibility for
these activities is that program administrators
presumably have in-depth knowledge of their programs and
an interest in improving educational substance, but they
may also have a vested interest in current operations.
At the same time, the central unit is likely to have less
program expertise but a greater concentration of
evaluation talent and social science expertise. When
such talent and expertise can be found to an adequate
extent in a program office, it may take the lead, with
the central evaluation unit as the cooperating office.
The central unit should also, from time to time, run
checks on accountability information developed by program
offices and the Inspector General and, when necessary,
conduct its own studies. Precisely how all these
evaluation responsibilities are shared between the
central unit and program offices ought to be a function
of the expertise residing in each program office.

Three functions are appropriately shared between the
central unit and NIE (which is under the Assistant
Secretary for Educational Research and Improvement). The
first is cost-benefit studies designed to establish the
efficiency of alternative ways of obtaining the
objectives of a given program. Such studies require all
the expertise needed for assessing program effects and
tying them to specific components of the program
treatment. In addition, benefits and costs of the
program must be put in monetary terms, a difficult
conceptual problem. Cooperation with NIE is suggested
because of the breadth of skills required and because it
may be necessary to conduct basic research in how to do
cost-benefit studies in education. Each particular
instance of doing such a study will provide material for
theoretical research and should be fully informed by it.
The two units should also jointly administer the types of
grant programs suggested in Chapter 3 for local and state

Beyond reporting, however, the central unit should be responsible for coordination of evaluation throughout the Department. Coordination is critical because of the interorganizational complexities discussed in Chapter 3. Many different parties within the Department have a stake in evaluation, most especially the operating program units and the planning component, which is currently separated from the central evaluation unit. (See the discussion below on the placement of the central evaluation unit.) Coordination also should contribute to more efficient use of evaluation resources. For the four phases of evaluation--planning, design of specific studies and procurement mechanisms, review, and use of findings--there are several ways in which authority and control could be distributed, i.e., in which evaluation activities could be coordinated:

1. The head of the central evaluation unit or cognizant assistant secretary could have both the responsibility and the authority (that is, final sign-off power) for approving plans, design and procurement, findings, and their dissemination. Insofar as possible, this person (office) could also set up incentives for application of findings or sanctions against nonuse.

2. The central unit could have major responsibility for coordination of planning, for review of designs and quality of procurement (but no sign-off power), and for review of findings together with the initiating unit, with dissemination also shared with that program unit.

3. Besides carrying out its own projects, the central unit could provide technical assistance (when asked) to other units engaged in evaluation activities, but have no further authority or responsibility. In this case, coordination responsibility or authority would either be assigned to some other level (say, the Secretary's or Undersecretary's office) or not assigned at all, as was the case for the Education Division within HEW until recently. (While HEW's Assistant Secretary for Planning and Evaluation received evaluation plans from the whole Education Division, generally only those from the central unit were reviewed; see Appendix A.)

The Committee believes that, for each phase of evaluatio . a different degree of sharing of responsibility and authority is appropriate. Relationships should also vary depending on the nature of the evaluation activity and the degree of expertise

residing in offices other than the central unit. We make some suggestions below as to coordinating mechanisms that strike a balance between totally centralized decision making (option 1 above) and autonomy for each unit (option 3 above). But we recognize that any (or no) coordination comes at a cost. The costs of no coordination at all include not only the wasteful use of evaluation dollars, but also the failure to use evaluation findings and the inability to cumulate knowledge about programs. The cost of any degree of coordination is time--more staff time for communication and more executive time for making decisions. Therefore, no matter what coordinative mechanisms are adopted, the Committee suggests that both the time invested and the results be tracked with some care, so that the effort to use evaluation resources wisely does not end up leading to negative results. For example, staff may get so occupied with meetings, with defenses against criticisms, and with waiting for decisions that they have inadequate time to produce procurement requests of high quality, to effectively monitor evaluation studies, to respond to modification requests from contractors or grantees, to review reports in detail, or to disseminate findings. Tracking of how well coordination procedures work should lead to their reexamination periodically, perhaps every 3 or 5 years. The rest of this section presents our suggestions for the Department with regard to coordination at each stage of the evaluation process.

Planning

We believe planning should be centralized, with all units--program, policy and planning, budget, research, etc.--involved at the staff level and with sign-offs required by each assistant secretary. The assistant secretary responsible for evaluation should take the lead for the coordination of planning. The central unit should carry responsibility for developing, together with the cognizant program units, a coordinated plan, including series of related studies, for each of the large federal education programs, as exemplified by those for Title I and P.L. 94-142. The central unit also should be charged with the coordination of all evaluation planning, even though the planning and execution of specific studies may be carried out elsewhere--a program office, the research unit, or even the local or state level.

We note the current attempt by the central evaluation
unit to coordinate plans for fiscal 1981 and fiscal 1982
(see U.S. Department of Education 1980b, 1980c). We
suggest coordination of planning not because we believe
that control of all evaluation activities should be
lodged in the central evaluation unit--we do not--but
because there appears to be no overall evaluation
planning with established goals and priorities for the
Department. Until the Department develops such plans, it
will be subject to ad hoc, arbitrary changes in
direction. Such changes prevent the cumulation of
incremental program information of the kind needed by
decision makers both in Congress and within the
Department. Improved evaluation planning will clarify
data and information needs for evaluation and allow the
Secretary to assign priorities to them in the context of
other data gathering needs. Recommendation D-10, which
speaks to this issue, is repeated here:

Recommendation D-10. The Department of Education should
institute a flexible planning system for evaluations of
federal education programs.

In Chapter 3 we emphasize that planning for evaluation
cannot be a totally internal activity. Outside groups
having a stake in a program must be consulted. Since the
Department's top priority external audience is Congress,
the Department needs to develop better liaison regarding
evaluation activities with members and with congressional
staff. Congressional aides have been very critical about
the relevance, timeliness, and packaging of evaluation
reports (see Zweig 1979). More involvement of
congressional staff is needed in selecting basic issues
and questions that can be answered by the evaluation
process. The central evaluation unit, being more removed
than program administrators from the politics surrounding
particular education programs, should be charged with the
responsibility of communicating with Congress about
evaluation needs (see Recommendation C-1). Program
units, on the other hand, tend to be closer to such
constituency groups as representatives of target
populations and educators charged with carrying out the
programs; therefore, they should be responsible for
obtaining their participation in the planning for
individual studies as well as in the development of the
overall plan.

Design of Studies and Procurement

Technical committees drawn from the staff of the central
evaluation unit and from the Office of Educational
Research and Improvement, supplemented by staff from the
originating unit (if other than either of these two)
should review and comment on all design and procurement
documents. Final veto or sign-off power, however, should
not reside with these committees but with the cognizant
assistant secretary supervising the unit that prepared
the design or the procurement instruments or grant
guidelines. If technical or substantive criticisms are
made by the reviewing committee, the cognizant assistant
secretary should require responses from the originating
unit that either refute the criticisms or indicate
changes made as a result. If the central unit is the
sponsor of the study, the process should be reversed,
with the relevant program unit providing review. The
central unit should also have staff available to provide
technical assistance during the execution of a study,
that is, when staff from other units monitoring an
evaluation contract might call for assistance in
reviewing progress or authorizing changes in study
direction, design, test instruments, analytical
strategies, and the like.

Review of Findings

The process for review of findings, either at an interim
stage or in final reports, should be similar to that
suggested for the design and procurement of studies.
Technical committees drawn from the staff of the central
evaluation unit and the Assistant Secretary for
Educational Research and Improvement (possibly the same
ones involved in the design and procurement phase) should
review reports and associated materials. Comments should
be forwarded to the originating unit, with a requirement
for rebuttal or incorporation of changes responsive to
the technical review. Program units should be afforded
the same review opportunity for studies originating in
the central unit. These internal reviews of designs and
of findings should be preliminary to the external reviews
suggested for each of these phases in Chapter 3.

Dissemination and Use

As recommended in Chapter 4, the originating unit should
have the responsibility of building a dissemination and
use plan into its original procurement document and of
ensuring that such a plan is part of the accepted
proposal and subsequent contract or grant. The
originating unit's dissemination plan would be reviewed
along with other features in the design and procurement
phase. The originating unit should have the
responsibility for carrying out the dissemination plan
addressed to the primary audiences, who presumably are
closely tied to the originating unit. The central unit
may carry out dissemination to secondary audiences as it
deems appropriate.

The central unit should also serve as the storehouse
and coordinating center for information derived from all
evaluation activities, including not only studies
originating in the Department, but also those carried out
by state and local agencies and even work relevant to
education that may not have been federally funded or be
concerned with federal education programs. The unit
should be responsible for cumulating knowledge from these
sources, reanalyzing data, and refocusing information
necessary to suggest changes in legislation, in
regulation, in program management, or in program
intervention as evidence indicates. Other units,
particularly the Department's research arm, should
cooperate in this integrative function.

Functioning as something like a nerve center for
evaluation information, the central unit should also be
charged with getting relevant information to audiences
that can act on it or are likely to have an interest in
it, beyond the audiences already included in the
dissemination plans for a specific study, as noted in the
following recommendations from Chapter 4:

Recommendation D-13. The Department of Education should
ensure that dissemination of evaluation results achieves
adequate coverage.

Recommendation D-14. The Department of Education should
observe the rights of any parties at interest and the
public in general to information generated about public
programs.

Recommendation D-15. The Department of Education should
give attention to the identification of "right-to-know"
user audiences and develop strategies to attend to their
information needs.

To carry out these functions, the central evaluation
unit should have a dissemination arm. Such a subunit
should also devote time and energy to the communications
problem. Too many evaluation reports are cloaked in
jargon that is unintelligible to decision makers and
other nontechnical audiences. Although most evaluation
contracts now specify that an executive summary must
accompany the final report, insufficient attention to
effective packaging of evaluation findings continues to
be the rule. Too many reports are not read or not
understood by busy policy makers or by outside groups
that could use the information because the language of
the reports is unclear. There is a real difference
between ambiguity of findings, which can be expected for
large, complex programs that encourage local variability,
and the inability to present those findings in
understandable prose. Personnel in the central unit
charged with responsibilities for disseminating
evaluation findings must perform the translation from
scientific jargon to clear English when such translation
has not been adequately done by contractors or grantees.
In order to be effective in this role, however, central
unit dissemination staff must possess requisite
communication skills and must be insulated from political
pressures that otherwise will quickly undermine the
credibility of their work.

Location of the Central Evaluation Unit

We have proposed that the central evaluation unit be
charged with important coordinating responsibilities in
developing the Department's overall evaluation plan and
in synthesizing and disseminating evaluation-related
knowledge derived from all sources. We do not foresee
that these responsibilities can be adequately carried out
as long as the central evaluation unit is subsumed within
the management arm of the Department. The implicit
message of this arrangement is that only the management
perspective of evaluation is considered a high priority.
While some members of the Committee favor the
assignment of an assistant secretary to the evaluation

function and other members disagree with this particular
approach, all members agree that evaluation is currently
too far removed from the top policy circles in the
Department. This distance makes it unlikely that the
central unit would be able to effectively coordinate
evaluation activities across the Department. Yet this
unit is probably the only one that could provide the
Secretary with a comprehensive view of the amount of
money being spent for evaluation, of the types of
evaluations under way, of the effectiveness of the
various disparate parts of the evaluation "system," and
of the potential for using study findings to make more
informed decisions about programs.

   A variety of administrative mechanisms can be used to
improve the current situation. For example, the
Department could make the unit a separate office
immediately responsible to the Secretary or the
Undersecretary to provide the needed access and
credibility. A precedent exists in the case of the
Office of Bilingual Education and Minority Languages
Affairs. Another possibility for making the unit more
effective is to couple it more closely to the major
planning function. We would caution, however, that some
separation should be maintained between evaluation and
budgeting. Though these functions are often located
together, subservience of evaluation to the budgetary
process is as counterproductive as using evaluation to
chastise or reward individual program managers,
apparently the Department's current direction. If
budgetary decisions and the handing out of rewards or
sanctions are to be the main functions of evaluation
activities, they will be devalued as a means for program
improvement. As long as evaluation is seen as a
threatening rather than as a supportive activity, those
who are subject to the threat will find ways of defusing
it by covert lack of cooperation or outright opposition.
As a result, evaluation activities will continue to be
curtailed, and results consigned to the dusty shelves of
unused reports.

                    CONSTRAINTS

No matter how evaluation responsibilities are assigned
and organized, the Department has to face some important
constraints that are only partly under its control:
constraints of budget, of staff, and of process.

                                    171

Budget Constraints

Pressures to reduce the federal budget have taken their
toll of evaluation projects since many such projects are
discretionary items. In fiscal 1980, the central
evaluation unit was not able to initiate any new studies
except those expressly mandated in law or made possible
through specific set-asides for evaluation (for example,
the half-percent of program funds mandated for national
evaluation of Title I). However, as a consequence of the
dispersion of evaluation responsibilities, the central
unit spends less than half the money invested in
evaluation at the national level: $19.6 million of the
$43.4 million estimated for the whole Department
(including the inspector general) in 1980. (For an
estimate of evaluation spending by various components of
the Department, see Appendix A.) As already noted,
additional federal funds are spent at the state and local
levels for evaluations. With respect to accountability
of spending for evaluation, then, there is trifurcation
of responsibilities: the central evaluation unit,
program units of the Department, and states and
localities. But only the central unit has been the
object of major scrutiny and a decreasing budget, while
responsibilities and funds are idiosyncratically assigned
by legislation or executive practice to selected federal
program offices and to state and local authorities, often
without similar scrutiny of performance.

   In the last 3 years, the Department has not been
successful in convincing the appropriations committees of
Congress that an increased budget for the central
evaluation unit was warranted, even while authorization
committees have asked for more evaluation. In fact,
funds have been appropriated for evaluation activities
outside the central unit, and Congress has spent
additional funds on its specially commissioned studies.
These actions appear to reflect an inability to make a
convincing case for the work of the central unit,
although it is not clear whether the apparent
dissatisfaction leading to decreasing budgets has been
warranted by inadequate performance or has been due to
greater visibility and scrutiny.

Staff Constraints

We have commented previously that the complexity of any
evaluation process beyond tracing money and counting

people calls for particular technical skills and social
science knowledge. Staff members responsible for
evaluation programs should be well grounded in the theory
and technical knowledge of a variety of social and
behavioral science disciplines. They must also be in
touch with the perspectives represented by various
interest groups who represent program beneficiaries and
service providers. Having practical program knowledge
and experience is helpful as well, though this can be
supplied through cooperation of the relevant program
units.

The staffs of evaluation offices have to be able to
explain issues involved, to develop questions to be
answered, to suggest methodologies for research, and to
prepare statements of work for RFPs and other procurement
documents. They have to participate in panels that
establish criteria and make recommendations for the
selection of winning contractors. They are also likely
to negotiate substantive contract issues before awards
are made. After a contract is awarded, the cognizant
staff person or project monitor must be able to provide
technical assistance if needed by the contractor, assist
in clearing survey instruments, and rule on modifications
requested by the contractor. In order to respond
effectively to contractor requests, the staff person
needs to understand through first-hand research
experience whether requested changes are appropriate or
not. Throughout the course of a project, staff members
must provide professional review, including careful
examination of final reports.

The unusual array of skills, experience, and diverse
perspectives needed to manage evaluation programs is not
easily obtainable. The Department is limited in its
ability to recruit top-quality staff in adequate numbers
because of personnel ceilings and other civil service
constraints. The Committee has not had time or
opportunity to assess the qualifications of the staff in
the central evaluation unit, though there are obvious
gaps in disciplinary expertise, in the representation of
minorities (see Chapter 3), and in hands-on experience
with field-based applied research studies of the kind
being designed and monitored by the unit. What seems
clear, however, is that the current deployment of staff
and assignment of responsibilities does not take
advantage of the collective expertise in the central unit
and in the research components located elsewhere in the
Department (for example, in NIE or the National Institute

173

of Handicapped Research). External requirements and
internal practice with respect to planning, procurement,
and clearance have severely constrained the time needed
to do quality work; the combined effect of
conceptualization of large-scale studies by single
individuals or small groups (as has been the practice in
the central unit) and the need for early closure on
technical detail is to leave little room for creativity.
Nor is it likely that the expertise represented by the
central unit is duplicated in every program office with
evaluation responsibilities. In some cases, evaluation
work carried out elsewhere in the Department may open up
innovative ways of planning and designing studies, as has
been true for the NIE compensatory education study and
the evaluation plan for P.L. 94-142. Both these
instances come from units with research expertise. Other
program offices, however, are unlikely to be able to
staff up for the evaluation responsibilities now assigned
them or that they might acquire in the future.

Recommendation D-17. The Department of Education should
examine staff deployment and should establish training
opportunities for federal staff responsible for
evaluation activities or for implementation of evaluation
findings.

The Department should consider alternative ways of
using the technical staff within the central unit and
evaluation staff in other units. Duties and
responsibilities would vary according to the amount of
government control exercised by staff: grants and
consultancies entail the least control; contracts and
evaluation teams configured of government staff and
outside experts more, and in-house studies the most.
Figure 5, adapted from one originally prepared by Wargo
(1980), illustrates the three major relationships between
government staff and outside experts and some of the
characteristics of each alternative. The Department has
largely used the contracting method, though in-house
analysis has been characteristic of selected areas,
particularly for postsecondary programs. There may be
evaluation work that is better addressed by the
grant/consultantship method (see Chapter 3) or by an
evaluation team. In part, the choice depends on the type
of evaluation work to be undertaken, but staff capability
is an equally important criterion. The greater the
degree of government involvement, the greater the skills

174

Government/Expert Relationship	Evaluation Role		Qualifications of Contractor/Panel Selection	Government Technical Expertise	Government Resources (time and staff)	Government Control	Contractor/Grantee Freedom	Probability of Timely Product
	Government	Grantee/Contractor/Consultant						
Grant/Consultantship (National Science Foundation) (National Institutes of Health)	• Defines problem area • Solicits proposals for area • External peer review panel selects grantee • Monitors proposal implementation	• Develops objectives, methods, and plan for evaluation • Conducts the evaluation with little guidance or technical support from government • Produces a final report without government input	High	Low	Low	Low	Low	
Contract (U.S. Department of Health and Human Services) (U.S. Department of Education)	• Defines evaluation objectives • Process oriented statement-of-work • Internal review panel selects contractor • Monitors all technical activities and expenditures • Some guidance, technical assistance, and direction provided	• Develops methods and implementation plan for evaluation • Receives some direction and technical support from government • Produces a final report that is reviewed and approved by government	Moderate	Low to moderate	Moderate	Moderate	Moderate	Moderate
Evaluation Team (Food and Nutrition Service of the U.S. Department of Agriculture)	• Specifies evaluation objectives and develops evaluation design • Design, process and management-oriented statement-of-work • In-house technical and program experts select contractor • Organizes, directs, and controls evaluation team consisting of evaluation and program staff, advisory panel, consultants, contractors, and subcontractors	• Critiques evaluation design and statement-of-work and suggests improvements • Becomes a key member of evaluation team • Participates in design, data collection, data analysis, and reporting • Receives day-to-day direction from government team director • All plans and products are developed as a team effort	High	High	High	High	Low	High
In-house Evaluation	• Substantially all work done by government staff	• Occasional reviews • Small contracts for data collection or analysis • Commissioned papers	—	Very high	Very high	Very high	—	High

FIGURE 5 Government evaluation relationships with experts.

and the greater the number of personnel that are
required. The grant/consultancy method allows maximum
contribution from the field; the evaluation team concept,
though it requires skilled staff, still allows
participation by outside experts while making possible
quick response (see Recommendation D-11).

For any given staff role in evaluation work, there
must be an adequate number of staff, and they must have
the requisite training and experience. Moreover, a work
atmosphere conducive to attracting good staff and holding
them must be provided. The Department should examine the
number and types of positions assigned to evaluation
activities in light of responsibilities and work load
(number of RFPs to be prepared, contracts monitored,
final reports to be analyzed, etc.) within the central
evaluation unit and wherever else evaluation activities
are carried out. It should also examine the extraneous
and counterproductive demands that are imposed on staff
through internal procedures that could be simplified.
Consideration of personnel needs should also take into
account the time required for the type of training
suggested below.

The academic and experience background of personnel
charged with evaluation responsibilities should be
examined in connection with the tasks they are required
to perform. This applies to staff in program units as
well as to staff in the central evaluation unit. If
necessary, training programs should be conducted to
prepare staff members for the writing of work statements,
to familiarize them with new evaluation techniques, and
to strengthen their knowledge of selected social science
disciplines. Handbooks should be prepared for persons
who monitor the substantive aspects of evaluation
contracts. If federal personnel lack field experience, a
determined effort should be made to expose them to
practical situations affecting the evaluation process.
Short-term field assignments could be used to provide
national office personnel with needed practical
experience.

At the same time, as noted in Recommendation D-4,
program executives and staff as well as other line
executives outside the units specifically concerned with
evaluation would benefit from greater knowledge of the
language of evaluation and how evaluations can be used.
Program managers at the federal level play a variety of
important roles in the evaluation of education programs.
Program managers often suggest which of the national

programs within their purview ought to be evaluated.
Such decisions reflect a concern for important issues in
program delivery and program effects that must be
translated into the evaluation questions to be asked.
Program managers need to provide key questions to the
evaluation experts, spell out what they consider to be
indicators of successful performance, and so on. During
the course of a study, managers often assume the role of
co-monitor and may accompany the technical evaluation
team into the field to assess progress. At the end of an
evaluation, managers play an important role in the
interpretation of the results. All of these roles would
be significantly improved if managers had a better
understanding of the basic principles of evaluation.
Training for federal staff on relevant topics should be
instituted. Seminars in evaluation methodology and in
applications of social science research to program
improvement could be given by technical staff from the
central evaluation unit and the Department's research arm
and by external evaluation experts. A newly created
training unit within the Department, the Horace Mann
Institute, provides an appropriate internal vehicle.
Other alternatives include specially tailored offerings
by the Federal Executive Institute and the Graduate
School of the Department of Agriculture (which is
scheduled for transfer to the Department of Education).
In addition to providing some technical knowledge,
training should increase the understanding of program
managers about what kind of information evaluation can
and cannot provide.

Process Constraints

In a number of ways, the Department's own procedures
inhibit its ability to produce timely and relevant
evaluation studies of high quality. These procedures
affect each stage of the process: producing a coherent
set of plans for the whole Department, designing
individual studies, procurement, launching the study once
a contract or grant has been awarded, monitoring its
progress, and disseminating its findings. A typical time
chart for a relatively straightforward study that is
intended to take 12 working months for design, data
collection, and analysis is pictured in Figure 6; under
current conditions, a lead time of 3 years is necessary.

Total anticipated
working time: 12 months
Contract duration: 18 months
Total elapsed time: 3 years

Award   Completion   Release
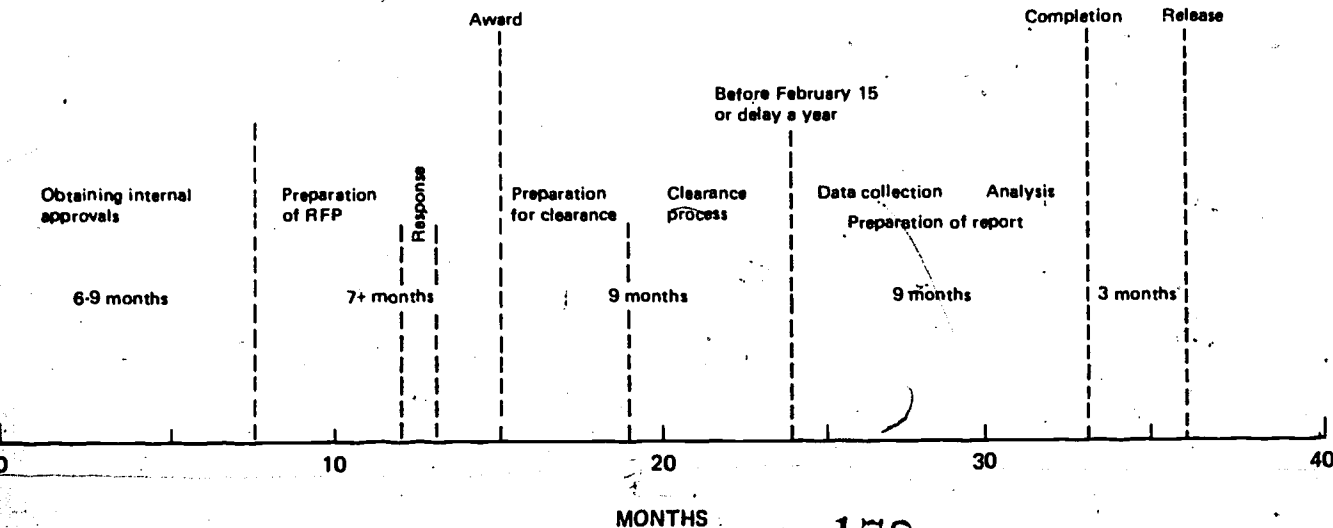
Before February 15
or delay a year

Obtaining internal   Preparation   Response   Preparation   Clearance   Data collection   Analysis
approvals   of RFP   for clearance process
                                                                                       Preparation of report

6-9 months   7+ months   9 months   9 months   3 months

0   10   20   30   40

MONTHS

178

FIG   Time chart -- typical evaluaton study.

Planning and Design

In recent years, the Department and its predecessors have
tried to introduce planning mechanisms that would help
set priorities and achieve greater coordination (see
Appendix A). One unfortunate consequence has been to
delay approval of studies, as illustrated by the 1980
procurement schedule (see Chapter 3, Table 1). Delays in
the planning process may create postponement of studies
into a new fiscal year. An even more adverse effect
(also noted in Chapter 3), has been the unwarranted
compression of time for the most difficult intellectual
work: design of a study by federal staff and by
responding proposers. The planning process is under the
control of the Department; presumably, as planning
mechanisms become better established, time delays can be
reduced.

Procurement

The procurement process or any alternative mechanism for
getting the work done entails negotiations within the
Department between the unit designing the evaluation and
the relevant program unit (if the study is not conducted
there) as well as other parties at interest, for example,
the Office of Civil Rights, the offices of the
Undersecretary or the Secretary, the Assistant Secretary
for Planning and Budget, or the National Institute of
Education. In selected cases--for example, in Title I
evaluations in which the legally constituted advisory
council participates--outside groups are also involved.
(We note that our recommendations in Chapter 3 with
respect to opening up the procurement process in order to
enhance the quality of evaluations will further
complicate the process and may introduce additional time
lags.) A major party to such negotiations is the Grant
and Procurement Management Division, which must approve
all procurement instruments or grants announcements. The
federal competitive procurement processes as interpreted
and enforced by this division take, on the average, 6
months from review of the statement of work prepared by
the initiating office to the time of award, exclusive of
response time allowed between announcement of RFP or
grant guidelines and the proposal due date.
Noncompetitive processes, such as sole-source awards or
unsolicited proposals, can be completed in shorter time,

but they are seldom employed because they are more easily
subject to the charge of favoritism.

The objectives of competitive procurement are to get
the best buy for the evaluation dollar and to assure a
fair process.[5] As the competitive procurement
mechanism now operates, neither objective is likely to be
attair.: only a few performers are able to compete, and
  .: costs of evaluations are increased by the
  siderable--though hidden--costs of the process
  reparing RFPs, writing lengthy proposals) that are
  :uilt into internal staff salaries and the total costs of
the resulting contracts. At the same time, the losses
that result from the process are considerable:
limitations on creativity and quality, time delays, and
wasteful use of human resources inside and outside
government. Though the way the government obtains
research services is generally regulated by statutes that
pose external constraints, any federal agency has
considerable latitude in its interpretation of applicable
regulations. Differences in operating procedures are
readily discernible to individuals familiar with several
agencies. The Department of Education would profit by
examining the more flexible strategies of other agencies.

Launching a Study

For any study that involves collecting the same
information from nine or more respondents, OMB clearance
(which may be delegated) must be obtained. When this
requirement was first instituted by OMB, there were three
reasons: to assure adherence to statistical standards,
to allow OMB to judge the economic impact of a proposed
study, and--most importantly in recent years--to reduce
the burden on respondents imposed by the multiplying
demands for data. Reduction of the response burden
remains a major objective for both the administration and
Congress. As more and more data collection efforts in
education became subject to clearance (e.g., program
report forms, statistics gathered by NCES, all evaluation
and research studies resulting in information to be
delivered to the government), the Education Division
within HEW set up its own internal screening mechanism,
the Educational Data Acquisition Council (EDAC), to
facilitate OMB clearance. In parallel, the chief state
school officers, concerned with the time and money
consumed by responding to federal data requests, also

obtained the right to clear study designs and instruments through their Committee on Education Information Systems (CEIS). The 1978 education amendments (P.L. 95-561) created the Federal Educational Data Acquisition Council (FEDAC, the successor to EDAC) as the designated body to replace OMB in controlling demand for data in education, with CEIS as an official participant. As noted in Chapter 3, the 1978 amendments also introduced the requirement for notification and availability by February 15 of data collection instruments to be used in the following school year. The effects of the clearance provisions are illustrated by the following examples.

A contract for a study on sex equity in vocational education, mandated by Congress, was awarded in late July 1977.[6] By early December, with concentrated efforts by the contractor and the federal project officer, the forms clearance package was sent to the OE clearance officer who had the job of reviewing submissions to EDAC. The clearance officer sent the package forward 2 months later, in early February 1978. EDAC clearance was obtained on March 1, and the package was then forwarded to the Assistant Secretary of Education whose clearance was needed before submission to OMB. This clearance was obtained on March 22, and OMB clearance, the final hurdle, received on April 14. Because the study had high visibility and because there were relatively few instruments involved, clearance took 4-1/2 months, close to the minimum time averaged during that period. There were, however, important changes in instrumentation: a major questionnaire dealing with attitudes was eliminated at the stage of OMB clearance (as were most such items in other types of instruments). The ostensible reason for the deletion was that the legislation did not require collection of that type of information. In this way, a review of 3 weeks overrode the work of 4 months--which included extensive consultation with parties at interest--by the contractor and the project monitor.

Another example concerns a planned study of Indian education scheduled for completion in order to feed into the reauthorization process for the legislation, due to expire in 1983; hence, the study results should be available for hearings likely to be held in 1982. Approval for the study was not received from within the Department until May 1980; an award was made on September 30, 1980. Even more than for the sex equity study, the choice and design of instrumentation will have to include careful consideration of the sometimes conflicting

sensitivities and points of views of the populations
being served, the service providers, and the framers of
the program--both legislative and executive. But the
previously noted requirement for February 15 notification
and availability of instruments means that there were
then only 4 months available to flesh out the design of
the study, including methods, measures, and specifics of
data collection, and for getting the whole package
approved through the clearance mechanisms. If the
February 15 deadline cannot be met, either a waiver will
have to be obtained or the study postponed for a whole
year. Not only will postponement add considerably to its
cost, but it will make the study irrelevant to the
purpose for which it is being undertaken, since data
collection could not even begin before fall of the year
(1982) in which the congressional hearings are to be held.

In another case, a recent 12-month study of OE
evaluation projects, clearance procedures had not been
completed by the time the study was done and the contract
had ended. The choice was to delete the data collection
aspect of the study or to proceed in the absence of
required clearance. The first would have led to a year
or more delay in the study, the second to illegal
procedures.

Carter (1977) describes two other examples. For the
sustaining effects study of Title I, a very complex study
using 10 different types of measures, clearance of the
first 2 of the 10 sets of measures took 8 months. The
clearance packages for all 10 sets of instruments
totalled 1,412 pages; Carter's estimate of the cost for
the clearance process (not including development of the
instruments) was $155,500 (in 1976 dollars). The second
example involved a congressionally mandated study of
Title I services for neglected or delinquent children;
clearance took 6 months. Carter notes (1977:11):

> Almost without exception the reviewers, either at
> OE or OMB, had never been to an institution for the
> neglected or delinquent. Many of them were not
> aware of the results of our clinical pretests, yet
> they felt they knew how and in what form the
> material should be collected. Again,
> office-generated expertise superceded actual field
> experience.

The last example that we cite provides an interesting
illustration of how the drive toward reducing

respondents' burden has created a lack of information in an area that would appear to be directly relevant to the federal role in education. Larson (1980) recently studied the collection of race, ethnic, and gender data on participants in federal education programs. Collection of such data is rare except in those cases were specific populations are targeted, for example in ESAA (desegregation assistance) and bilingual programs, for which information on the targeted group is collected. Other exceptions are research studies not directly coupled to specific program evaluations, such as the National Assessment of Educational Progress (1978). Yet given the overall mission of the federal education programs to increase equal educational opportunity, it is somewhat surprising that programs as a whole are not evaluated with respect to their effectiveness in improving education for ethnic or racial minorities and females. Recently, regulations have been changed to make possible the gathering of data on race, ethnicity, and gender in grantee applications for funds, but the gathering of such data for program assessment has always been possible. That it is still largely absent can in good·part be ascribed to budgetary and clearance constraints, which drive any evaluation study toward collecting only those data for which there is an express "need-to-know." And "need-to-know" is often equated with specific mention of a subgroup in legislation for the program or its evaluation.[7] One can only conclude that current clearance procedures, whatever other purpose they may serve, have had the effect of minimizing the ability to obtain information crucial to meeting federal goals in education. In part, that effect may have been the result of considering each study in isolation as it went through the clearance process and attempting to minimize response burden case by cas' We note that the process is in the midst of change.

At this time, t . intent (expressed both through executive action by OMB· and through proposed legislation in Congress) is to manage the reduction of response burden more like fiscal budget allocations: each agency submits to OMB an information collection budget that requests an allocation of the total number of burden hours necessary to carry out its management, evaluation, and research responsibilities. On the basis of the submission, an allocation will be made by OMB, probably with a 10-15 percent cut in response burden, a goal announced for 1981. (Another cut is to be made the

following ye⋅ for a total cut of 30 percent over 2
years.) The agency will then reallocate the information
collection budget internally. In the case of the
Department of Education, of 8-1/2 million burden hours
that were requested, some 7 million hours, or more than
80 percent, is allocated to program administration and
compliance, that is, information to be submitted by
program applicants and grantees, information needed for
fiscal audits, and information needed to enforce
compliance with civil rights laws. OMB will delegate the
responsibility for clearance of specific studies and
instruments to the agency's internal mechanism when it is
deemed to be functioning well or the law so specifies, as
is the case with FEDAC.[8]
    The evolution of clearance procedures from reviewing
individual studies to a process that assembles all
proposed data collection in one document should allow
top-level Department officials to consider the data needs
of evaluation and research in a forum where they are
presented together with those of program administration,
enforcement (for example, the data needs of the Office of
Civil Rights), auditing, and the periodic gathering of
general statistical data and indicators (for example, the
data collected by NCES). It may also encourage the
coordination of studies across organizational units so
that studies proposed by one unit can use data collected
elsewhere. The Department should be alert to the
opportunities for more coherent evaluation and data
collection activities offered by the new clearance
process.

Progress

After clearance, time delays in the progress of a study
will be occasioned by the inevitable discrepar⋅ es
between assumptions in the study design and a⋅ ⋅⋅
conditions in the field. The nature of the p⋅⋅⋅⋅⋅
activity, the individuals engaged in it, the w⋅⋅⋅⋅⋅⋅⋅
of respondents to cooperate, the presence of
documentation--all will present unforeseen difficulties,
particularly if the timing of the study is thrown off
schedule by the clearance process. Other delays may be
introduced by the researchers themselves, who are wary of
potential criticism and therefore employ time-consuming
procedures to assure technical impeccability that does
not enhance the quality of the study (e.g., by meticulous

but unwarranted cleaning of data sets). Federal monitors
are often not in a position to know whether such
procedures are necessary or which delays in the progress
of the study are legitimate. Sanctions against
contractors who do not deliver products on schedule are
seldom enforced since extenuating circumstances can
always be cited. This is particularly true because of
the inability of federal monitors to respond in timely
fashion to simple, much less to complex, requests for
changes in the study plan, either because of their
workload or because they do not have authority on their
own to rule on the requested change. Hence, delay
becomes no one's responsibility.

Dissemination

Within HEW in recent years, dissemination of study
findings has been held up in the Secretary's office for
many montns because of the perceived need to have the
Secretary informed and able to respond to inquiries from
the media and the public. For example, for the study on
sex equity in vocational education (referred to above)
the findings were not released until nearly a year after
the final report was submitted in April 1979. The delay
appeared to be occasioned by the controversial subject of
the study rather than by the findings themselves, since
no changes were made in the final report (Harrison and
Dahl 1979).

The advent of the new Department of Education brought
about new rules: a directive on release of findings
(U.S. Department of Education 1980a) provides 10 days,
after acceptance of the study report by the central
evaluation unit, for response from program and other
offices. Reports are to be released after the 10-day
period, accompanied by the comments received. However,
this rule does not deal with delays occasioned by
disagreements between the sponsoring office and the
performers or with release of findings by sponsoring
offices other than the evaluation unit. For example, one
study report submitted in January 1980 (David 1980),
whose findings were in dispute between the sponsoring
office (the Assistant Secretary for Program Evaluation in
the former HEW) and the then central evaluation unit for
the Office of Education, had still not been released 10
months later. Congress in particular i 3 been concerned
with such delays: on occasion, the suspicion has arisen

that findings were not being released because they did
not support the positions of the current administration
with respect to the program that was the subject of the
study.

In summary, process constraints have become severe in
recent years. It is not unlikely that, during the time
it takes to complete a study, conditions in the field or
policy concerns regarding a specific program will change,
making the findings of the evaluation, when they do
become available, of little interest.

Recommendation C-5. Congress should authorize a study
group to analyze the combined effects of the legislative
provisions and executive regulations that control
federally funded applied research.

Congress has been dissatisfied with the lack of
relevance and timeliness of much evaluation work in
education. One of the causes for delay and for
irrelevance is the accumulation of rules and regulations
governing the relationships between sponsor, researcher,
and action site or agency, i.e., the Department of
Education, the contractor, and the state/school/student.
The whole process of funding and carrying out applied
research about social services is severely constrained by
these rules and by the operating precedents they have
engendered. Almost every provision now on the books or
enforced through executive practice may be justified when
considered in isolation: to prevent favoritism in
contract awards, to protect respondents from a heavy
burden of requests for data, to protect the privacy of
individuals, to require disclosure of information related
to the public business, and so forth. Their combined
effect, however, has been to lengthen the time needed for
compliance, to increase the costs both within government
(through greater investment of staff time) and of
extramural contracts and grants, and to discourage whole
classes of potential performers from participating.
Though laws sometimes specify time limits for procedures
(e.g., for OMB clearance of data collection instruments),
they are seldom observed in practice.

To date, most of the concern has been with instituting
procedures to guard against possible transgressions in
initiating and carrying out applied social science
research. The trade-offs between the benefits of such
safeguards and the obstacles they create to producing

timely and relevant applied research at reasonable cost
have been largely ignored. It is not clear how much of
the negative effect is due to the laws and regulations
themselves and how much to the interpretation and
operational mechanisms within any given agency. For this
reason, the recommended analysis must go beyond the
problems within a single agency or department and examine
the process as it works in several different agencies.

Recommendation D-18. The Department of Education should
take steps to simplify procedures for procuring
evaluation studies, carrying them out, and disseminating
their findings.

   The Committee has recommended (see Chapter 3) that the
means by which the Department solicits, selects, and
funds evaluation studies be expanded in order to allow
more performers to participate. The competitive
procurement process involving issuance of an RFP and
awarding of a contract to the highest-ranked or
lowest-priced bidder is by far the most commonly used
form of solicitation. This type of solicitation was
designed by the government for the purchase of highly
specifiable goods or services so that contracts could be
awarded on the basis of the best buy for the dollar. The
rules that have accumulated over the years to ensure fair
competition have shifted considerable control of the
process from the technical specialists (for example, in
the evaluation unit or in a research office) to the
contracting office, the interpreters and enforcers of the
government procurement regulations. This has had serious
implications for the quality of evaluations (discussed in
Chapter 3) and has increased the time needed for arriving
at compromises acceptable to all. The process has become
not only restrictive and inflexible but very costly in
internal staff time and for potential contractors. And
since the cost to contractors is recouped eventually from
the government through overhead and in other ways, the
government bears the double burden.
   Recent criticisms (U.S. General Accounting Office
1980a, Gup and Neumann 1980) have focused on abuses
possible in the use of consultants and sole-source
contracting. The Committee is not convinced that the
cost of rules instituted to prevent such abuses is not
higher than the cost of the abuses themselves. The
various means (other than competitive procurement through

RFPs) that can be used to obtain evaluation work are
discussed in Chapter 3 (see Recommendation D-5). The
Department must be more deliberative in choosing whether
to use competitive procurement, sole-source contracting,
8-A contracting, cooperative agreements, basic ordering
agreements, or grant awards, within the limitations of
the law (see P.L. 95-224).

The major sources of delay, once a contract or grant
for a study has been awarded, must also be identified and
addressed. This applies particularly to clearance
procedures and to the in-house handling of requests for
changes in study design, sampling procedures, testing,
analysis, time frames, and the like. While a request for
a modification is being considered, the evaluation may be
in a hold status, pending the sponsor's response. In
such cases, the sponsor's nonresponsiveness can
contribute materially to delays in project completion,
with the effect of cost overruns.

At times, failure to perform on time is the
responsibility of the contractor or grantee. The
Department should institute and enforce sanctions and
incentives to encourage timely performance. For example,
some agencies have included clauses in contracts that
provide that nontimely performance (products not
delivered by the specified date) can be a basis for
nonpayment of up to one-third of the contractor's fee.

Most contracted evaluations have provisions for review
of delivered products by the project officer, which often
may entail extensive internal review and clearance. To
the extent that these reviews are not completed in an
efficient and timely manner, the projects are subjected
to time delays. Such delays may be as injurious as
budget overruns, leading to delays in dissemination of
findings and charges of lack of timeliness. Because of
the possible cost of such delays, Recommendation D-13
(see Chapter 4) seeks to limit the period of control over
evaluation results. The Committee is not advising
against review: quite the contrary. It is advocating
that the time taken for internal review be shortened in
favor of making findings freely available to stand the
test of the marketplace. In the long run, this will both
increase the quality and improve the chances of
appropriate use of evaluation results.

193

NOTES

1   There are exceptions. Political appointees given the
    job of reducing the budget will have reasons to find
    reduced needs.
2   At present, the Office of Civil Rights (OCR) is
    funding a study to review testing and evaluation
    instruments used with handicapped persons and another
    study to identify the factors that cause
    overrepresentation of minority children in programs
    for the mentally retarded. OCR has also funded
    cost-benefit analyses of programs mandated under civil
    rights legislation (O'Neill 1976).
3   For fiscal 1980, the budget for the Department of
    Education was $14.2 billion. For ESEA Title I, the
    1980 budget provided $3.2 billion; for Education for
    the Handicapped, $1.05 billion, and for Rehabilitation
    Services and Handicapped Research $932 million; for
    vocational education, $928 million; for impact aid,
    $825 million; for emergency school aid, $249 million;
    and for bilingual education, $167 million.
4   As an example, when the National Institute of
    Education was under an edict from its governing body,
    the National Council for Educational Research, to
    increase the percentage of funds spent for basic
    research, it shifted its labeling of certain
    activities from "evaluation" to "research." Since the
    boundaries are often fuzzy, this kind of redefinition
    is not infrequent. As a counterexample, nearly $1
    million allocated to the evaluation of Title VII
    (bilingual education) were reprogrammed in fiscal 1980
    by the former Assistant Secretary for Education in HEW
    to support further development of "Villa Alegre" (the
    bilingual analog to "Sesame Street"), a decrease of
    more than one-third in the actual evaluation budget,
    though reporting figures stayed unchanged (see
    Appendix A).
5   "Best" has different connotations in different
    instances: it may mean the lowest-priced proposal of
    those technically acceptable; it may mean the
    lowest-priced proposal of those exhibiting high
    degrees of excellence; or it may mean some combination
    of these and other criteria spelled out in the RFP.
6   The information on this study was provided by Robert
    Maroney and Dorothy Shuler of the Office of Program
    Evaluation, the central evaluation unit. Their help
    in tracing the clearance procedures and other process
    steps is gratefully acknowledged.

7 In her study, Larson gives some additional reasons for
   the absence of data on race and sex variables: such
   data are deemed to be irrelevant or dangerous, they
   raise costs by requiring larger samples, and they are
   the concern of enforcement rather than of evaluation
   staff.

8 FEDAC has a permanent staff of four professionals,
   augmented by three to four professionals on detail
   from other units or from outside the Department. From
   time to time, however, FEDAC staff are themselves
   detailed for considerable periods of time to other
   duties. Staff shortage has been a major cause of
   delays in obtaining clearance.

Glossary

AERA	American Educational Research Association
AFT	American Federation of Teachers
AIR	American Institutes for Research
ARROE	American Registry of Research and Related Organizations in Education
'ASPE	Assistant Secretary for Planning and Evaluation
ASPIRA	An educational research group oriented toward Puerto Rican interests
BEH	Bureau of Education for the Handicapped (OE), now Division for Special Education and Rehabilitation Services
BOAE	Bureau of Occupational and Adult Education (OE), now Division of Vocational and Adult Education
CCSO	Council of Chief State School Officers
CEIS	Committee on Evaluation and Information Systems
CENTRAL (EVALUATION) UNIT	formerly the Office of Program Planning, Budgeting, and Evaluation (OPPBE/OE), later the Office of Evaluation and Dissemination (OED/OE), now the Office of Program Evaluation (OPE/ED)

179

COSHMHO   formerly the Coalition of Spanish-Speaking
Mental Health Organizations, now the
National Coalition of Hispanic Mental
Health and Human Services Organizations

CRS   Congressional Research Service

DISTAR   A reading program for primary grades

DOL   U.S. Department of Labor

ED   U.S. Department of Education

EDAC   Educational Data Acquisition Council

EDUCOM Inc   A private corporation performing
educational research and development

ERIC   Educational Resources Information Center

ESAA   Emergency School Assistance Act

ESAA-TV   A series of television programs aimed at
minority group children of school age

ESEA   Elementary and Secondary Education Act

FEDAC   Federal Educational Data Acquisition Council

FNS   Food and Nutrition Service of the U.S.
Department of Agriculture

FY   Fiscal Year

GAO   U.S. General Accounting Office

GPMD   Grant and Procurement Management Division
(OE)

HEW   U.S. Department of Health, Education, and
Welfare

HHS   U.S. Department of Health and Human Services

IDEA   Institute for Development of Educational
Activities

IEP	Individual Education Plan (P.L. 94-142)
IG	Inspector General
ISA	Intermediate Service Agency (set up by SEAs and LFAs to provide services to LEAs)
ISD	Independent School District
JDRP	Joint Dissemination Review Panel (OE-NIE)
LEA	Local Education Agency
MDRC	Manpower Demonstration Research Corporation
NAACP	National Association for the Advancement of Colored People
NCES	National Center for Education Statistics
NEA	National Education Association
NIE	National Institute of Education
NIH	National Institutes of Health
NSF	National Science Foundation
OE	Office of Education
OED	Office of Evaluation and Dissemination (central evaluation unit in OE)
OMB	Office of Management and Budget
OPE	Office of Program Evaluation (current designation of central evaluation unit in Division of Management, ED)
OPPBE	Office of Program Planning, Budgeting, and Evaluation (former title of central evaluation unit in OE)
PAC	Parent Advisory Committee (Title I, ESEA)
P.L.	Public Law (for example, P.L. 94-142, Education for All Handicapped Children Act of 1975)

193

R&D	Research and Development
RDD&E	Research, Development, Dissemination, and Evaluation
RDU	Research and Development Utilization Program (NIE)
RFP	Request for Proposal
RFQ	Request for Qualifications
SBA	Small Business Administration
SDC	Systems Development Corporation
SEA	State Education Agency
TAC	Technical Assistance Center (Title I, ESEA)
USOE	U.S. Office of Education

References

Abramson, M.A. (1978.) The Funding of Social Knowledge
   Production and Application: A Survey of Federal
   Agencies. Washington, D.C.: National Academy of
   Sciences.

Abt, C.C. (1979) Government constraints on evaluation
   quality. In L.E. Datta and R. Perloff, eds.,
   Improving Evaluations. Beverly Hills, Calif.: Sage
   Publications.

Acland, H. (1975) Secondary Analysis of the Emergency
   School Assistance Program. R-1969-HEW. Santa Monica,
   Calif.: Rand Corporation.

Alkin, M.C., Daillak, R., and White, P. (1979) Using
   Evaluations: Does Evaluation Make a Difference?
   Beverly Hills, Calif.: Sage Publications.

Baker, E.L., ed. (1980) Evaluating Federal Education
   Programs. Report No. 153. Los Angeles, Calif.:
   Center for the Study of Evaluation, University of
   California.

Baker, K., and Ginsburg, A. (1980) A Rat in a Penny
   Arcade: Some Observations on Policy Relevance and the
   Sustaining Effects Study. Paper presented at the
   annual meeting of the American Education Research
   Association, Boston.

Ball, S., and Bogatz, G.A. (1970) The First Year of
   "Sesame Street." Princeton, N. J.: Educational
   Testing Service.

Berman, P., and McLaughlin, M.W. (1975-78) Federal
   Programs Supporting Educational Change, Vols. I-VIII.
   R-1589/1-HEW to R-1589-8-HEW. Santa Monica, Calif.:
   Rand Corporation.

183

Bernstein, I.N., and Freeman, H.E. (1975) Academic and Entrepreneurial Research: The Consequences of Diversity in Federal Evaluation Studies. New York: Russell Sage Foundation.

Berryman, S.E., and Glennan, T.K., Jr. (1980) Ai improved strategy for evaluating federal programs in education. Chapter 2 in J. Pincus, ed., Education Evaluation in the Public Policy Setting. Santa Monica, Calif.: Rand Corporation.

Bogatz, G.A., and Ball, S. (1971) The Second Year of "Sesame Street": A Continuing Evaluation. Princeton, N.J.: Educational Testing Service.

Boruch, R.F., and Cordray, D.S. (1980) An Apprai l of Educational Program Evaluations: Federal, State, and Local Agencies. Prepared for U.S. Department of Education. Northwestern University.

Brickell, H.M. (1974) Data for Decisions: An Analysis of Evaluation Data Needed by Decision Makers in Educational Programs. New York: Policy Studies in Education.

Campbell, D.T., and Erlebacher, A.E. (1970) How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth, ed., The Disadvantaged Child, Vol. 3. New York: Brunner/Mazel.

Caplan, N. (1976) The use of social statistics by federal executives with special attention to policy decisions in education. Appendix E in National Research Council, Setting Statistical Priorities. Washington, D.C.: National Academy of Sciences.

Caplan, N. (1977) A minimal set of conditions necessary for the utilization of social knowledge in policy formulation at the national level. Pp. 183-197(b) in C.H. Weiss, ed., Using Social Research in Public Policy Making. Lexington, Mass.: D.C. Heath.

Caplan, N. (1980) Knowledge Conversion and Utilization. Paper presented at conference on "The Political Realization of Social Science Knowledge and Research: Toward New Scenarios." Institute for Advanced Studies, Vienna, Austria.

Caplan, N., Morrison, A., and Stambaugh, R.J. (1975) The Use of Social Science Knowledge in Policy Decisions at the National Level: A Report to Respondents. Ann Arbor, Mich.: Institute for Social Research, University of Michigan.

Carter, L.F. (1977) Federal clearance of educational evaluation instruments: Procedural problems and proposed remedies. Educational Researcher 6(5):7-12.

Chelimsky, E., ed. (1977) An Analysis of the Proceedings of a Symposium on the Use of Evaluation by Federal Agencies. Report M77-39, Volumes 1 and 2. McLean, Va.: MITRE Corp.--Metrek Division.

Chelimsky, E. (1978) A Framework for Balancing Privacy and Accountability Needs in Evaluations of Social Research. Letter of December 11 to Associate Director of Program Analysis Division, General Accounting Office.

Chen, H.-T., and Rossi, P.H. (1980) The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science. Social Forces 59(September):106-122.

Cicirelli, V., and Granger, R. (1969) The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development. PB 184328. Westinghouse Learning Corporation, Ohio University. Distributed by Clearinghouse for Federal Scientific and Technical Information, Springfield, Va.

Cohen, M.D., and March, J.G. (1974) Leadership and Ambiguity. New York: McGraw-Hill.

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

Coleman, J.S., Bartot, V., Lewin-Epstein, N., and Olson, L. (1979) Policy Issues and Research Design. Report to the National Center for Educational Statistics (Contract No. 300-78-0208). National Opinion Research Center, University of Chicago.

Coleman, J.S., Campbell, E.Q., Hobsen, C.J., McPartland, J., Mood, A., Weinfeld, F.D., and York, R.L. (1966) Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office.

Conant, J.B. (1959) The American High School Today. New York: McGraw-Hill.

Cook, T.D., Appleton, H., Conner, R.F., Shaffer, A., Tamkin, G., and Weber, S.J. (1975) "Sesame Street" Revisited. New York: Russell Sage Foundation.

Cook, T.D., and Gruder, C. (1979) Metaevaluation research. Chapter 24 in L. Sechrest, S.G. West, M.A. Phillips, R. Redner, and W. Yeaton, eds., Evaluation Studies Review Annual, Vol. 4. Beverly Hills, Calif.: Sage Publications.

Crain, R.L., and York, R.L. (1976) Evaluating a successful program: Experimental method and academic bias. School Review 84(2):233-254.

Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess,
   R.D., Hornik, R.C., Phillips, D.C., Walker, D.F., and
   Weiner, S.S. (1980) Toward Reform of Program
   Evaluation: Aims, Methods, and Institutional
   Arrangements. San Francisco, Calif.: Jossey-Bass,
   Inc.

Cross, C.T. (1979) Title I evaluation--a case study in
   congressional frustration. Educational Evaluation and
   Policy Analysis 1(2):15-21.

Danoff, M.N. (1978) Evaluation of the Impact of ESEA
   Title VII Spanish/English Bilingual Education
   Program--Overview of Study and Findings. Palo Alto,
   Calif., American Institutes for Research.

Datta, L. (1975) Design of the Head Start planned
   variation experiment. Pp. 79-87 in A.M. Rivlin and
   P.M. Timpane, eds., Planned Variations in Education.
   Washington, D.C.: Brookings Institution.

Datta, L. (1976) The impact of the Westinghouse/Ohio
   evaluation on the development of Project Head Start.
   Pp. 129-191(b) in C.C. Abt, ed., The Evaluation of
   Social Programs. Beverly Hills, Calif.: Sage
   Publications.

Datta, L. (1979) Damn the Experts and Full Speed Ahead:
   An Examination of the Study of Federal Programs
   Supporting Educational Change as Evidence Against
   Directed Development and for Local Problem Solving.
   National Institute of Education, Washington, D.C.

David, J.L. (1978) Local Uses of Title I Evaluations.
   Research Report EPRC 21. SRI Project URU-6854. Menlo
   Park, Calif.: SRI International.

David, J.L. (1980) Federally Mandated Evaluation
   Policy: The Evolution of the Title I Evaluation
   Program. Unpublished paper submitted to Education
   Planning Division, U.S. Department of Health,
   Education, and Welfare, Washington, D.C.

Davis, H.R. (1973) Change and innovation. In S.
   Feldman, ed., Administration and Mental Health.
   Springfield, Ill.: Charles C. Thomas.

Dearman, N.B., and Plisko, V.W. (1979) The Condition of
   Education. Statistical report of the National Center
   for Education Statistics. Washington, D.C.: U.S.
   Government Printing Office.

Diaz, W.A., Ball, J., Jacobs, N., Solnick, L, and
   Widman, A. (1980) The Youth Entitlement
   Demonstration: Second Interim Report on Program
   Implementation. New York: Manpower Demonstration
   Research Corporation.

Dresek, M., and Higgins, P.H. (1980) Current vs. Ideal
Procedures for Determining Educational Program
Evaluation Budgets: A Survey of School Evaluators.
Paper presented at Division H Task Force meeting of
the American Educational Research Association annual
meeting, Boston.

Elmore, R.F. (1975) Design of the Follow Through
experiment. In A.M. Rivlin and P.M. Timpane, eds.,
Planned Variations in Education. Washington, D.C.:
Brookings Institution.

Evaluation Research Society (1980) Standards for Program
Evaluation. Exposure Draft. Evaluation Research
Society, Potomac, Md.

Ezrahi, Y. (1978) Political contexts of science
indicators. Chapter 12 in Y. Elkana, J. Lederberg,
R.K. Merton, A. Thackray, and H. Zuckerman, eds.,
Toward a Metric of Science: The Advent of Science
Indicators. New York: John Wiley and Sons.

Far West Laboratory for Educational Research and
Development (1979) Educational Programs That Work,
6th ed. Prepared for the National Diffusion Network
of the U.S. Department of Education. Available from
Far West Laboratory for Educational Research and
Development, 1855 Folsom Street, San Francisco, Calif.

Florio, D.H. (1980) Educational Inquiry and the
Congress: Unmet Potential or Undervalued Resource?
Paper presented at the annual meeting of the American
Educational Research Association, Boston.

Glaser, E.M. (1973) Knowledge transfer and institutional
change. Professional Psychology 4:434-444.

Gramlich, E.M., and Koshel, P.D. (1975) Educational
Performance Contracting. Washington, D.C.: Brookings
Institution.

Gregg, G., Preston, T., Geist, A., and Caplan, N. (1979)
The caravan rolls on: Forty years of social problem
research. Knowledge: Creation, Diffusion,
Utilization 1(1):31-61.

Gup, T., and Neumann, J. (1980) Government out of
control: Contracts. Washington Post (June 23-26,
1980).

Harrison, L., and Dahl, P.R. (1979) Sex Equity Study of
Vocational Education. Palo Alto, Calif.: American
Institutes for Research.

Havelock, R.G. (1969) Planning for Innovation Through
Dissemination and Utilization of Knowledge. Ann
Arbor, Mich.: Institute for Social Research,
University of Michigan.

Havelock, R.G., and Lingwoood, D.A. (1973) R&D
  Utilisation Strategies and Functions: An Analytical
  Comparison of Four Systems. Ann Arbor, Mich.:
  Institute for Social Research, University of Michigan.
Hedrick, T.E., Boruch, R.F., and Ross, J. (1979) On
  ensuring the availability of evaluative data for
  secondary analysis. Chapter 19 in L. Sechrest, S.G.
  West, M.A. Phillips, R. Redner, and W. Yeaton, eds.,
  Evaluation Studies Review Annual, Vol. 4. Beverly
  Hills, Calif.: Sage Publications.
Hill, P.T. (1980) Evaluating education programs for
  federal policymakers: Lessons from the NIE
  compensatory education study. Chapter 4 in J. Pincus,
  ed., Educational Evaluation in the Public Policy
  Setting. Santa Monica, Calif.: Rand Corporation.
HOPE Associates, Inc. (1979) Performance Review of
  USOE's ESEA Title I Evaluation Technical Assistance
  Program. HOPE Associates, Inc., Washington, D.C.
  (Helping Other People Everywhere)
Human Interaction Research Institute (1976) Putting
  Knowledge to Use: A Distillation of the Literature
  Regarding Knowledge Transfer and Change. Joint
  project with National Institute of Mental Health. Los
  Angeles, Calif.: Human Interaction Research Institute
Joint Committee on Standards for Educational Evaluation
  (1980) Standards for Evaluation of Educational
  Programs, Projects, and Materials. New York:
  Harcourt, Brace.
Katz, E., and Lazarsfeld, P.F. (1955) Personal
  Influence: The Part Played by People in the Flow of
  Mass Communications. New York: Free Press.
Kirst, M., and Jung, R. (1980) The utility of a
  longitudinal approach in assessing implementation: A
  thirteen-year view of Title I, ESEA. Educational
  Evaluation and Policy Analysis 2(5):17-34.
Larson, M. (1980) An Exploration of the Use of Race and
  Sex as Variables in Federally Sponsored Education
  Evaluations. Draft submitted to National Institute of
  Education, Department of Health, Education, and
  Welfare, Washington, D.C.
Levitan, S.A., and Wurzburg, G.K. (1979) Evaluating
  Federal Social Programs: An Uncertain Art.
  Kalamazoo, Mich.: W.E. Upjohn Institute for
  Employment Research.
Lippitt, R.O. (1965) The use of social research to
  improve social practice. American Journal of
  Orthopsychiatry 35(4):663-669.

Logan, C.H., (1972) Evaluation research in crime and
    delinquency: A reappraisal. _Journal of Criminal Law,_
    Criminology and Police Science 63:378-387.
Lyon, C.D., Doscher, L., McGranahan, P., and Williams,
    R. (1978) _Evaluation and School Districts._ Los
    Angeles, Calif.: Center for the Study of Evaluation,
    University of California.
Manpower Demonstration Research Corporation (1979)
    Summary of the Operating Experience and Statistical
    Highlights of the Third Year of the National Supported
    Work Demonstration. New York: Manpower Demonstration
    Research Corporation.
Marsh, J.C., Geist, A., and Caplan, N. (in press) _Legal_
    Reforms in the Prevention and Control of Rape.
    Boston: Auburn House.
Martin, R., and McClure P. (1969) Title I of ESEA: Is
    It Helping Poor Children? Washington Research Project
    of the Southern Center for Studies in Public Policy
    and the NAACP Legal Defense of Education Fund, Inc.
Maynard, R.A., Brown, R., and Schore, J. (1979) _The_
    National Supported Work Demonstration: Effects During
    the First 18 Months After Enrollment. New York:
    Manpower Demonstration Research Corporation.
McLaughlin, M. (1975) _Evaluation and Reform: The Case_
    of ESEA, Title I. Cambridge, Mass.: Ballinger.
Minnesota Systems Research, Inc. (1976) _Final Report:_
    Evaluation and Prediction of Methodological Adequacy
    of Research and Evaluation Studies. Minneapolis,
    Minn.: Minnesota Systems Research, Inc.
Mosher, F.C. (1979) _The GAO: The Quest for_
    Accountability in American Government. Boulder,
    Colo.: Westview Press.
Mosteller, F., and Moynihan, D.P, eds. (1972) _On_
    Equality of Educational Opportunity. New York:
    Vintage.
Murphy, J.T. (1974) _State Education Agencies and_
    Discretionary Funds. Lexington, Mass.: D.C. Heath.
National Advisory Council on Equality of Educational
    Opportunity (1975) _Third Interim Report._ Washington,
    D.C.: U.S. Government Printing Office.
National Assessment of Educational Progress (1978)
    Series of reports on national assessments of changes
    in achievement, 1969-1977. National Assessment of
    Educational Progress, Denver, Colo.
National Institute of Education (1976) _Evaluating_
    Compensatory Education: An Interim Report on the NIE
    Compensatory Education Study. Washington, D.C.:
    National Institute of Education.

National Institute of Education (1977) Reports on the
Compensatory Education Study: Administration of
Compensatory Education; Compensatory Education
Services; Demonstration Studies of Funds Allocation
Within School Districts; The Effects of Services on
Student Development; Title I Funds Allocation: The
Current Formula; Using Achievement Test Scores to
Allocate Title I Funds. Washington, D.C.: National
Institute of Education.
National Opinion Research Center (1973) Southern
Schools: An Evaluation of the Effects of the
Emergency School Assistance Program and of School
Desegregation. NORC 124A. National Opinion Research
Center, University of Chicago.
National Science Foundation (1979) Federal Funds for
Research and Development: Fiscal Years 1977, 1978,
and 1979, Volume XXVII. NSF 79-310. Washington,
D.C.: U.S. Government Printing Office.
Office of Child Development (1973) OCD--Head Start
Policy Manual. Washington, D.C.: U.S. Department of
Health, Education, and Welfare.
Office of Management and Budget (1973) Social
Indicators, 1973. Washington, D.C.: U.S. Government
Printing Office.
Office of Management and Budget (1977) Resources for
Program Evaluation, Fiscal Year 1977. Washington,
D.C.: U.S. Office of Management and Budget.
Office of Management and Budget (1979) Management
improvement and the use of evaluation in the executive
branch. Circular No. A-117. Office of Management and
Budget, Washington, D.C.
O'Neill, D.M. (1976) Discrimination Against Handicapped
Persons. Arlington, Va.: Public Research Institute.
Patton, M.Q., Grimes, P.S., Guthrie, K.M., Brennan, N.J.,
French, B.D., and Blyth, D.A. (1977) In search of
impact: An analysis of the utilization of federal
health evaluation research. Pp. 141-163 in C.H.
Weiss, ed., Using Social Research in Public Policy
Making. Lexington, Mass.: Lexington Books.
Raizen, S.A. (1979) Dissemination programs at the
National Institute of Education: 1974 to 1979.
Knowledge: Creation, Diffusion, Utilization
1(2):259-292.
Reisner, E.R. (1980) The Office of Education Administers
Changes in a Law: Agency Response to Title I, ESEA
Amendments of 1978. Durham, N.C.: NTS Research
Corporation.

Resmovic, H.L. (1979) Methodological considerations in evaluating correctional effectiveness: Issues and chronic problems. Pp. 163-209 in L. Sechrest, S.O. White, and E.D. Brown, eds., The Rehabilitation of Criminal Offenders: Problems and Prospects. Washington, D.C.: National Academy of Sciences.

Robertson, L.S. (1980) Crash involvement of young teenaged drivers when driver education is eliminated from high school. American Journal of Public Health 70(6):599-603.

Rock, W.C. (1980) Personal communication.

Rogers, E.M. (1962) Diffusion of Innovations. New York: Free Press.

Rogers, E.M., and Shoemaker, F.F. (1971) Communication of Innovations: A Cross-Cultural Approach. New York: Free Press.

Rossi, P.H. (1979a) Issues in the evaluation of human services delivery. Chapter 3 in L. Sechrest, S.G. West, M.A. Phillips, R. Redner, and W. Yeaton, eds., Evaluation Studies Review Annual, Vol. 4. Beverly Hills, Calif.: Sage Publications.

Rossi, P.H. (1979b) Past, present, and future prospects for evaluation research. In L.E. Datta and R. Perloff, eds., Improving Evaluations. Beverly Hills, Calif.: Sage Publications.

Rossi, P.H., Freeman, H.E., and Wright, S.R. (1979) Evaluation: A Systematic Approach. Beverly Hills, Calif.: Sage Publications.

Schmidt, R.E., Scanlon, J.W., and Bell, J.B. (1979) Evaluability Assessment: Making Public Programs Work Better. Human Services Monograph Series, No. 14. Washington, D.C.: Project Share.

Schreier, M.A. (1979) Program participants' positive perceptions: Psychological conflict of interest in social program evaluation. Chapter 10 in L. Sechrest, S.G. West, M.A. Phillips, R. Redner, and W. Yeaton, eds., Evaluation Studies Review Annual, Vol. 4. Beverly Hills, Calif.: Sage Publications.

Sechrest, L. (1980) Evaluation researchers: Disciplinary training and identity. Pp. 1-18 in L. Sechrest, ed., Training Program Evaluators (Report #8 in New Directions for Program Evaluation). San Francisco, Calif.: Jossey-Bass, Inc.

Social Science Research Council (1978) Audits & Social Experiments: A Report Prepared for the U.S. General Accounting Office. Committee on Evaluation Research. PAD-79-1. Washington, D.C.: U.S. General Accounting Office.

Stanford Research International (1977a) Research on the
  Effectiveness of Compensatory Education Programs: A
  Reanalysis of Data. Final Report. Menlo Park,
  Calif.: Stanford Research International.
Stanford Research International (1977b) Trends in
  Management of ESEA Title I: A Perspective from
  Compliance Reviews. Vol. I: Overview, findings, and
  conclusions. Menlo Park, Calif.: Stanford Research
  International.
Systems Development Corporation (1976) Policy Questions
  Addressed by the Sustaining Effects Study. Santa
  Monica, Calif.: Systems Development Corporation.
Tallmadge, G.K. (1977) The Joint Dissemination Review
  Panel IDEABOOK. Prepared for the U.S. Department of
  Health, Education, and Welfare. Washington, D.C.:
  U.S. Department of Health, Education, and Welfare.
U.S. Congress (1977) Response to AIR Study. House
  hearings on HR 15, Testimony of the Center for Applied
  Linguistics, April 18, 1977. Washington, D.C.: U.S.
  Government Printing Office.
U.S. Department of Commerce (1977) Social Indicators,
  1976. Washington, D.C.: U.S. Government Printing
  Office.
U.S. Department of Education (1980a) Memoranda:
  Departmental Policy on the Distribution of Evaluation
  Executive Summaries and Study Reports. From the
  Deputy Assistant Secretary for Evaluation and Program
  Management-Designate (J. Seal).
U.S. Department of Education (1980b) Memoranda: Fiscal
  Years 1981 and 1982 Guidance for Evaluation
  Activities. From the Assistant Secretary for
  Management (J. Gabusi).
U.S. Department of Education (1980c) Memoranda:
  Recommendations for the FY 1981 and FY 1982 Evaluation
  Plan. From the Assistant Secretary for Management (J.
  Gabusi).
U.S. Department of Health, Education, and Welfare (n.d.)
  Evaluation of the Education for All Handicapped
  Children Act, Public Law 94-142. Division of
  Innovation and Development, Bureau for the Education
  of the Handicapped. Washington, D.C.: U.S.
  Department of Health, Education, and Welfare.
U.S. Department of Health, Education, and Welfare
  (1979a) Annual Evaluation Report on Programs
  Administered by the U.S. Office of Education--Fiscal
  Year 1979. Prepared by the Office of Education.
  Washington, D.C.: U.S. Department of Health,
  Education, and Welfare.

U.S. Department of Health, Education, and Welfare
(1979b) Briefing Book on Office of Evaluation and
Dissemination. Internal Office of Education
document. U.S. Department of Health, Education, and
Welfare, Washington, D.C.
U.S. Department of Health, Education, and Welfare
(1979c) Memorandum of December 21 from J.J. Vanecko.
Draft Report on Title I Evaluation Planning.
U.S. General Accounting Office (1977) Problems and
Needed Improvements in Evaluating Office of Education
Programs. HRD-76-165. Washington, D.C.: U.S.
General Accounting Office.
U.S. General Accounting Office (1978) Assessing Social
Program Impact Evaluations: A Checklist Approach.
PAD-79-1. Washington, D.C.: U.S. General Accounting
Office.
U.S. General Accounting Office (1979) Guidelines for
Model Evaluation. Exposure Draft. Washington, D.C.:
U.S. General Accounting Office.
U.S. General Accounting Office (1980a) Controls Over
Consulting Service Contracts at Federal Agencies Need
Tightening. PSAD-80-35. Washington, D.C.: U.S.
General Accounting Office.
U.S. General Accounting Office (1980b) Evaluating
Programs with Multiple and Conflicting Objectives: A
Framework and Some Methods. Exposure Draft. PAD-80.
Washington, D.C.: U.S. General Accounting Office.
Valentine, J., and Zigler, E., eds. (1979) Project Head
Start: A Legacy of the War on Poverty. New York:
Free Press.
Wargo, M.J. (1980) Federal Government's Evaluation
Relationships with Experts. Invited address to
Washington Operations Research and Management Council,
Washington, D.C., October.
Webster, W.J., and Stufflebeam, D.L. (1978) The State of
Theory and Practice in Educational Evaluation in Large
Urban School Districts. Paper presented at the annual
meeting of the American Education Research
Association, Toronto.
Weiss, C.H. (1977) Introduction. Pp. 1-22 in C.H.
Weiss, ed., Using Social Research in Public Policy
Making. Lexington, Mass.: Lexington Books.
Wholey, J.S. (1979) Evaluation: Promise and
Performance. Washington, D.C.: Urban Institute.
Wholey, J.S., Nay, J.N., Scanlon, J.W., and Schmidt,
R.E. (1975) Evaluation: When is it really needed?
Evaluation 2(2):89-93.

Wilson, J.Q. (1980) Round and round with consultants.
    Washington Post (July 11, 1980):A-13.
Wisler, C.E., and Anderson, J.K. (1979) Designing a
    Title I evaluation system to meet legislative
    requirements. Evaluation and Policy Analysis
    1(2):47-56.
Yin, R.K. (1980) Who Controls Evaluation? The
    Interorganizational Complexities of Evaluation
    Research. In E.L. Baker, ed., Evaluating Federal
    Education Programs. Report No. 153. Los Angeles,
    Calif.: Center for the Study of Evaluation,
    University of California.
Yin, R.K., Heald, K.A., Vogel, M.E., Fleischauer, P.D.,
    and Vladeck, B.C. (1976) A Review of Case Studies of
    Technological Innovations in State and Local
    Services. R-1870-NSF. Santa Monica, Calif.: Rand
    Corporation.
Zaltman, G., Duncan, R., and Holbek, J. (1973)
    Innovations and Organizations. New York: Wiley.
Zweig, F.M. (1979) The evaluation worldview of
    congressional staff. Pp. 145-158 in F.M. Zweig, ed.,
    Evaluation in Legislation. Beverly Hills, Calif.:
    Sage Publications.

A

Federal Evaluation Activities in Education: An Overview
Elizabeth R. Reisner

Federal funds support a broad range of program evaluation
activities in education. Such activities range from
national studies involving achievement testing of
thousands of students to local assessments of federally
supported projects in individual school districts.

   This paper is intended to provide an overview of those
federal evaluation activities that are designed to yield
information on federal education assistance programs.
The first section of this paper describes the major
evaluation activities of each of the organizational units
making up the former Education Division of the former
U.S. Department of Health, Education, and Welfare (HEW)
and certain other units. Taken together, these units
constitute the main offices currently conducting
evaluation activities in the U.S. Department of Education
(ED). Information on evaluation activities of these
offices is presented in tabular form and contains (1) a
listing of the major federal education programs being
evaluated by each of the organizational units sponsoring
education evaluations, (2) a description of each unit's
principal evaluation objectives, and (3) a rough estimate
of the fiscal 1980 funds used for evaluation by each of
the units.

The author is a senior policy analyst with NTS Research
Corporation in Washington, D.C. Previously, she had
staff responsibility for the review of evaluation
planning in the Office of Education.

195

The second section of this paper provides anecdotal
information on federally supported evaluations conducted
by state and local agencies. The third section describes
the evolution of the federal role in the evaluation of
education programs. The final section describes the
process used for deciding what national studies of
federal education programs are conducted and what
questions those studies address.

Information for this study was collected in interviews·
with federal managers whose offices are responsible for
conducting program evaluations in education as well as
from the works listed in the references. In several
instances internal memoranda of HEW, the Office of
Education (OE), and ED were used as source materials.
Because the intent of the paper is to present a broad
overview of the topic, it has been necessary to summarize
detailed information in a number of cases; the author
accepts full responsibility for any unintentional errors
of fact or emphasis that may have occurred in preparing
the summaries.

Authority for the Department of Education was enacted
on October 17, 1979, as P.L. 96-88, the Department of
Education Organization Act. The act permitted a 6-month
implementation period prior to official start-up of the
new department. ED was officially inaugurated on May 4,·
1980. In this paper, policies and procedures in effect
prior to that date are described using the earlier
organizational terminology (e.g., OE and the
Commissioner). Current terminology (e.g., ED and the
Secretary) is used to describe activities occurring after
May 4, 1980.

MAJOR EVALUATION ACTIVITIES
OF THE HEW EDUCATION DIVISION

Table A-1 provides summary descriptions of federally
supported evaluation activities designed to provide
information relevant to programs administered by the
former HEW Education Division. The primary offices
within the Education Division were OE, the National
Institute of Education (NIE), and the National Center for
Education Statistics (NCES); these offices are now
organizationally situated within ED. The information in
Table A-1 pertains primarily to former Education Division
offices because these are the offices for which
comparable information was most readily available.

Information from HEW's planning office and from the
office of the Inspector General is also included, along
with data on activities of the U.S. General Accounting
Office. Although data in the table were compiled in May
1980, there have not been major changes in the use of
fiscal 1980 funds.

A broad, inclusive definition of program evaluation
was used in compiling the data presented in Table A-1.
It is adapted from the definition used by Robert Boruch
in his proposal to OE to conduct a study of federally
supported education evaluations at state and local levels
(discussed in the second section of this report).
Boruch's definition, which is consonant with that used by
the Committee (see Chapter 2), includes the following
activities under the heading of program evaluation:
needs assessments, surveys, and other assessments
conducted prior to program initiation or review; process,
or formative, assessments intended to yield descriptive
information on the composition, organization, or
activities of a program; outcome, or summative,
assessments intended to yield information on the relative
benefits, costs, and other effects of a program; and
cost/benefit analyses intended to draw together
information on several types of program effects.

The category headings used in Table A-1 are as follows:

* "Federal office conducting evaluation activities"
refers to offices implementing evaluations (for in-house
efforts) and offices overseeing evaluation contracts (for
contracted studies). The organizational headings do not
necessarily reflect offices of equal bureaucratic rank.
* "Programs being evaluated" refers to the
principal federal programs that are being studied.
* "Main evaluation objectives" reflects the
priorities as described by federal evaluation managers in
interviews for this project and in written statements
prepared as part of the HEW evaluation planning process.
The information in the table does not include federally
supported evaluations conducted by local projects for
purposes of either self-assessment or fulfillment of
federal program requirements.
* "Federal funds used for evaluation in fiscal
1980" comprises estimates reported by evaluation managers
and described in internal planning papers. Funds used in
fiscal 1980 are indicated because that is the most recent
year for which fairly precise estimates are available.

TABLE A-1 Federal Evaluation Activities in Support of Programs Administered by the HEW Education Division

Federal Office Conducting Evaluation Activities	Programs Being Evaluated	Main Evaluation Objectives	Federal Funds Used for Evaluation in Fiscal 1980 ($ thousands)	Special Features
Office of Education (OE)				
Office of Evaluation and Dissemination (OED)				
Elementary and secondary programs	Title I of the Elementary and Secondary Education Act (ESEA), emergency school aid, bilingual education, Title IV civil rights, national diffusion network, and impact aid.	Assessment of impact of program services on students (e.g., Title I, bilingual, and emergency school aid); description of program services, especially with regard to beneficiaries (e.g., Title I) and classroom activities (e.g., bilingual); provision of technical assistance for the improvement of state and local evaluations (e.g., Title I).	15,400	Impact studies play a decreasing role in overall efforts; increasing emphasis on support to state and local evaluation activities and on measurement of federal program implementation at the state and local levels.
Occupational, handicapped, and developmental programs	Vocational education, education of the handicapped, adult education, Indian education, libraries, educational technologies, and special projects (e.g., teacher centers and basic skills).	Response to congressionally mandated studies (e.g., vocational education, career education, and community education); information on impact of service delivery programs (e.g., libraries); exploratory evaluations, as described at entry for HEW Assistant Secretary for Planning and Evaluation (e.g., gifted and talented).	2,500	Significant portions of overall funding come from regular program accounts and from program administrative accounts, at the decision of program managers (e.g., Indian education and community education).
Postsecondary programs	Postsecondary grant and loan programs for students and discretionary grant programs for institutions (e.g., developing institutions and special services for disadvantaged students).	In student aid, (1) assessment of program impact as measured by reduction of financial barriers for students and (2) improvement in management of aid programs; in institutional aid programs, assessment of impact in terms of (1) increased financial stability and	1,700	

210

program quality (e.g., developing
institutions) and (2) increased
enrollment rates of disadvantaged
students (e.g., special services
for disadvantaged students).

Subtotal for OED   19,600

Bureau of Education for the Handicapped (BEH)	Programs administered by the Bureau, especially state grants for education of handicapped children.	Fulfillment of mandated study and reporting objectives in Education of All Handicapped Children Act (P.L. 94-142), with special attention to the state approaches and practices that are most effective in the identification and delivery of services to handicapped children. Current projects include surveys of local student assessment practices and local services to handicapped children.	1,250	Evaluation activities based on multiyear evaluation plan developed and distributed following enactment of P.L. 94-142.
Bureau of Student Financial Assistance (BSFA)	Postsecondary grants to students (e.g., basic educational opportunity grants) and loans (e.g., guaranteed student loans).	Objectives similar to the student aid-related objectives of OED Postsecondary Programs Office, except less emphasis on program impact and more emphasis on management improvement; special attention to collection and analysis of data necessary for adjusting aid formulas to target intended students, while reducing instances of fraud and abuse of federal funds.	1,800	
Bureau of Occupational and Adult Education (BOAE)	Programs authorized by the Vocational Education Act and Adult Education Act.	Assessment of current needs for vocational education and technical assistance for state and local evaluations of federally supported activities.	500	Activities carried out primarily by National Center for Research in Vocational Education located at Ohio State University (total fiscal 1980 funding was $5.5 million).

199

211

TABLE A-1 (continued)

Federal Office Conducting Evaluation Activities	Programs Being Evaluated	Main Evaluation Objectives	Federal Funds Used for Evaluation in Fiscal 1980 ($ thousands)	Special Features
Follow Through	Follow Through, a discretionary grant program to aid disadvantaged children in the primary grades (K-4).	Current stress on improving the delivery of services to Follow Through grantees through the development of performance indicators for project implementation. Research, now separate from evaluation, concerned with development of new models and analysis of variables affecting implementation of models.	1,000 for evaluation and research combined (full 1,000 used to compute OE subtotal below).	Evaluation and research activities have been transferred from OED to program office, as a result of recommendation from exploratory evaluation conducted by Assistant Secretary for Planning and Evaluation.
Subtotal for OE			24,150	
National Institute of Education (NIE)				
Testing, assessment, and evaluation	Several small, urban education programs (e.g., push-excel and cities in schools) and general state and local instructional programs.	Improvement of local instructional practice, through (1) evaluations aimed at meeting needs identified by local instructional and administrative personnel (for small urban programs), (2) assistance to state and local educational personnel in improving quality of evaluations, and (3) research in evaluation methods.	4,000	Evaluation efforts not primarily oriented toward improvement in implementation of major federal education programs.
Mandated study of vocational education	Program administered by BOAE and, to a lesser extent, the Department of Labor's employment training programs for young people.	Assessment of policies (e.g., improvement in match between training activities and job opportunities) and mechanisms (e.g., open planning process) underlying the Vocational Education Act; studies not intended to evaluate current program impact on students.	1,000	Study mandated by Congress in the Education Amendments of 1976; comprehensive study plan submitted to Congress at beginning of study, followed by periodic reports. Overall effort patterned after congressionally mandated study of compensatory education 1975-78.

212

Dissemination and improvement of practice	NIE programs concerned with dissemination and improvement of practice, e.g., state capacity building for dissemination, RDU, ERIC, and the women and minorities program.	Assessment of four NIE programs to determine effectiveness of approaches to the transfer of educational research and development to educational practitioners.	625	Current NIE programs in area of knowledge transfer being used as vehicles for research into alternative methods of educational dissemination.
Subtotal for NIE			5,625	
National Center for Education Statistics (NCES)	In addition to regular educational surveys, special quasi-evaluation activities are as follows: ● Evaluation of NCES technical assistance to users of NCES data. ● Validity studies of ongoing surveys (Vocational Education Data System and Higher Education General Information Survey in FY 1980). ● Fast Response Survey on policy issues, as requested by policy offices.	Assessment of NCES's own effectiveness in (1) making its data accessible and relevant to users (e.g., evaluation of NCES technical assistance) and (2) designing and implementing its surveys (e.g., validity studies); also, provision of needs-assessment-type data on a rapid basis (3-6 months) for use in policy making.	855	All NCES activities have potential evaluation-related uses, since they may provide information on need for changes or adjustments in federal programs. Total appropriation for NCES in fiscal 1980 was $10 million.
HEW Office of the Assistant Secretary for Planning and Evaluation (ASPE)				
Education Planning	Education Division programs with large fiscal outlays (e.g., ESEA Title I) or with especially promising educational approaches (e.g., Fund for the Improvement of Postsecondary Education).	Examination of national policy alternatives, often through re-analyses of data collected by other agencies (e.g., OE, NCES, Bureau of the Census); oversight of OE evaluation activities (e.g., ESEA Title I).	300.	Occasional requests for OE to conduct specific evaluation studies.

TABLE A-1 (continued)

Federal Office Conducting Evaluation Activities	Programs Being Evaluated	Main Evaluation Objectives	Federal Funds Used for Evaluation in Fiscal 1980 ($ thousands)	Special Features
Evaluation	Representative sample of OE programs, broken down into "large formula grants," "large discretionary grants," and "small discretionary grants."	Identification of measurable program objectives and development of appropriate measures for use by program managers in assessing whether objectives are being met (e.g., Follow Through and bilingual education).	500	Exploratory evaluation approach being used for some studies conducted by OE.
Subtotal for ASPE			800	
HEW Office of the Inspector General (IG)	OE programs with large fiscal outlays (e.g., ESEA Title I, and post-secondary grant and loan programs) and programs with legislatively mandated audit requirements (e.g., vocational education state grants).	Auditing of activities at federal, state, and local levels to determine (1) adherence to principles of sound fiscal management and (2) compliance with pertinent legal requirements (e.g., Title I requirement that federal dollars must supplement and not supplant state and local spending on target children).	10,000 (200 staff-years, estimated at $50,000 per staff-year).	Planned HEW/IG activities for fiscal 1980 reportedly canceled in anticipation of new IG for Department of Education.

214

| General Accounting Office (GAO) | OE programs or program components believed to have serious management problems (e.g., developing institutions, student aid eligibility for proprietary institutions, and defaults in the guaranteed student loan program) or unclear program objectives (e.g., Follow Through and bilingual education); also programs coming up for reauthorization in Congress. | Assessment of the federal administration of educational programs and evaluation of program impact on intended beneficiaries. Studies focused on generating program recommendations for Congress and for relevant federal agencies. | 2,500 (50 staff-years, estimated at $50,000 per staff-year). | Programs selected for review according to requests from members of Congress or GAO staff. |

Subtotal for all offices except IG and GAO			31,430	
TOTAL			43,930	

- "Special features" contains miscellaneous
information relevant to evaluation activities of several
of the offices indicated.

Among the categories of information presented in Table
A-1, the category most vulnerable to change is the annual
funding data. These amounts are subject to considerable
fluctuation within any given year because of decisions to
move funds into or out of accounts previously designated
for evaluations and because of different interpretations
as to whether a given project is an evaluation or a
research activity. An example of the first type of
fluctuation was the decision early in fiscal 1980 to
transfer funds out of the "line item" appropriation for
studies and evaluation of bilingual education programs in
order to fund a bilingual television project. A total of
$700,000 in OE funds for federal program administration
was designated to be used to replace the transferred sum,
but because of high expenses associated with implementing
the new ED, the bilingual evaluation funds were not
replaced. An example of the second type of fluctuation
can be seen in NIE's reports of its own program
expenditures. Because of an administrative decision to
allot the maximum amount of NIE's funding to research
purposes, the Institute intentionally labels very few of
its projects as evaluations, even though many have
characteristics that conform to the definition presented
above.
    The aspect of the table most likely to provoke
questions from readers is the inclusion of federally
conducted audits of federal, state, and local
implementation of federal programs. Audits are generally
not considered to be evaluative in nature, especially
since they usually focus only on the fiscal operations of
individual federally funded projects. In recent years,
however, federal audits have become increasingly
concerned with nonfiscal matters, particularly state and
local compliance with legislated objectives and
procedures. The adoption of this auditing focus has
resulted, in some instances, in a blurring of the
distinction between audits and evaluations, particularly
given the establishment of specified national priorities
for federal education audits. For example, the fiscal
1980 work plan for the HEW Office of the Inspector
General identified three priorities for audits of state
and local administration of Title I of the Elementary and
Secondary Education Act (ESEA): (1) compliance with the

216

Title I statutory requirement for annual maintenance of
local fiscal effort per pupil; (2) implementation of
Title I state requirements for monitoring and enforcement
plans; and (3) operations of the centralized Migrant
Student Record Transfer Service funded under Title I.
With the establishment of explicit compliance-oriented
auditing objectives such as these, federally conducted
audits have acquired a distinct resemblance to program
evaluations.

                FEDERALLY SUPPORTED EVALUATIONS
              CONDUCTED BY STATE AND LOCAL AGENCIES

Virtually all federal education aid programs require
institutional grantees to conduct evaluations of their
own performance. The specific language of the
legislative requirements varies among programs, depending
on the overall objectives of the program and also on the
evaluation methodologies considered by federal
administrators to be best suited to the particular
program. For programs with a large state administrative
role, such as ESEA Title I and the state grant program
under the Vocational Education Act, states are also
required either (1) to collect local evaluation data and
provide summaries of these data to ED on a regular basis
or (2) to carry out their own state-managed evaluation
efforts.

    In recent years congressional mandates and Education
Division program managers have identified state and local
evaluaton priorities with increasing specificity, but the
offices of the former Education Division do not at
present collect regular data on the implementation of
state and local evaluation requirements. Therefore, it
is not possible to determine what portion of ED program
grant funds are used by grantees for self-evaluation
purposes nor is it possible to determine exactly how
those funds are used. It is apparent, however, that
significant amounts of federal funds are used to provide
assistance to state and local agencies in improving the
quality of their evaluations.

    Evaluations conducted by state and local agencies are
generally funded using program grant funds. At the state
level, evaluation activities are supported using state
administrative funding provided by the pertinent federal
program. ESEA Title I, for example, provides each state
educational agency with 1.5 percent of the state's total

Title I funding for purposes of state administrative
activities, including Title I program evaluation. In
school year 1979-80, amounts available for Title I state
administrative activities, including evaluation, ranged
from $4.5 million in New York to $225,000 in the 14
states with the lowest Title I enrollments. Other
federal education programs also provide administrative
funding to state education agencies.

At the local level, evaluation activities must be
supported out of each school district's federal grant
funds. The district's grant application usually
describes the evaluation activities planned by the
district and indicates how much of its grant is proposed
to be used for evaluation purposes. That proposal is not
generally binding on the district, however, once the
federal grant is received. (For more detail on the
funding and management of local evaluation activities,
see Appendix C.) Examples of the types of state and
local evaluation activities carried out under three
federal education programs are described below.

ESEA Title I

As a result of a requirement contained in the Education
Amendments of 1974 (P.L. 93-380), OE developed a set of
local evaluation models for use by Title I grantee. The
models, as specified in federal regulations (45 C 116.7
and 116a.50-57 published in the Federal Register
October 12, 1979), provide methods for measuring it
achievement gains in reading, mathematics, and lar ja
arts. ED (and formerly OE) also provides technical
assistance (at a cost of $11 million in fiscal 1980) to
state education agencies on methods for assisting local
districts in the use of the models. Despite extensive
efforts by OE since 1974 in designing and implementing
the models, Congress has expressed concern in committee
reports for the Education Amendments of 1978 (P.L.
95-561) that the Title I evaluation models do not yield
data that can be used by local Title I administrators as
a basis for improving Title I projects (U.S. Congress
1978a:51, and U.S. Congress 1978b:29-30). Findings in
support of this view have also been presented by David
(1980) and Orland (1980), but they are contradicted by
statements of the ED evaluation office.

ESEA Title VII

The-Education Amendments of 1974 also mandated that
evaluation models be developed for use by local districts
receiving funds under ESEA Title VII, the Bilingual
Education Act. The Education Division did not
immediately implement that mandate, however, and it was
reiterated in the Education Amendments of 1978. The
Senate committee report on the 1978 amendments expressed
hope "that these guidelines will provide scientifically
valid information as well as describe the unique features
of each project in order that local level projects can be
validly compared" (U.S. Congress 1978b:69). The ED
evaluation office is currently overseeing a project
intended to yield evaluation models for use by Title VII
grantees. In early descriptions of the project, the
evaluation office has stated that the models are to be
designed on the basis of existing approaches (including
the current Title 1 evaluation models) and are not to
reflect any new or "basic research."

As in Title I, the Title VII program also funds
technical assistance providers who are expected to assist
local Title VII grantees in improving the quality of
their self-evaluations. Until the evaluation models are
ready, however, grantees and assistance providers have
relatively little guidance on which to base local
evaluation efforts, except for criteria in the Title VII
final regulations requiring attention to "data collection
instruments and methods," "data analysis procedures,"
"time schedules," and the like (45 CFR 123a.30(e)
published in the Federal Register on April 4, 1980).

Vocational Education Act

The Education Amendments of 1976 (P.L. 94-482)
established a comparable set of requirements for the
Vocational Education Act. States are required to use
"statistically valid sampling techniques" to measure "the
extent to which program completers and leavers (i) find
employment in occupations related to their training, and
(ii) are considered by their employers to be well-trained
and prepared for employment" (Section 112 (b)(1)(B) of
the Vocational Education Act). In addition, the
legislatively mandated "national center for research in
vocational education" is to "work with states, local
educational agencies, and other public agencies in

developing methods of evaluating programs, including the
follow-up studies of program completers and leavers
required by Section 112, so that these agencies can offer
job training programs which are more closely related to
the types of jobs available in their communities,
regions, and states..." (Section 171(a)(2) of the Act).
The national center at Ohio State University has prepared
materials relevant to their technical assistance role; a
recent list of their activities includes three projects
aimed at implementing this mandate: "Evaluation Services
for Education Agencies," "Evaluation Handbooks," and
"Increasing the Credibility of Vocational Education
Evaluations" (listed in Gordon et al. 1979:62-63, 153).
The NIE mandated study of vocational education is
currently examining the performance of states in
implementing their evaluation requirements.

Studies of State and Local Evaluation Activities

Despite these extensive statutory mandates for state and
local evaluations, the only effort up to now to review
federally supported state and local evaluations across
federal programs has been the recent study by Boruch and
Cordray (1980). That study provides information on those
state and local evaluation activities aimed at producing
data relevant to federal categorical programs. There are
also three studies (one of which is under way now) that
provide information on state and local evaluation
activities supported from a variety of sources, federal
and nonfederal.

Survey of large school district evaluation units. The
Center for the Study of Evaluation at the University of
California at Los Angeles has examined the organization
of local school district offices of evaluation. This
survey acquired data on the size, staffing, and
organizational structure of evaluation offices in school
districts with enrollments over 10,000 (Lyon et al. 1978).

Survey of educational researchers and research
organizations. Under contract with NIE, the Bureau of
Social Science Research in 1976-78 surveyed nonfederal
organizations conducting research, development,
dissemination, and evaluation activities in education.
Information was obtained on funding, organizational
characteristics, and activities of 2,434 such entities
(Frankel et al. 1979) (see Appendix B).

Study of how school districts use information from testing and evaluation. Currently under way through an NIE contract to the Huron Institute in Cambridge, Massachusetts, this study is intended to develop strategies for helping school districts make better use of evaluation and test information. Initial reports from the study were made available in the fall of 1980; the final report is to be issued in the fall of 1981.

Although each of these studies sheds light on state and local evaluation activities in education, none provides a comprehensive description of state and local evaluations undertaken to assess the operations of federal programs.

EVOLUTION OF THE FEDERAL ROLE
IN THE EVALUATION OF EDUCATION PROGRAMS

Evaluation requirements are a relatively recent addition to federal education programs. The first mandatory evaluations for an OE program were those carried out by local districts implementing ESEA Title I projects. In 1965 Senator Robert Kennedy introduced language into the draft version of Title I requiring that "effective procedures, including provision for appropriate objective measurements of educational achievement, will be adopted for evaluating at least annually the effectiveness of the programs in meeting the special educational needs of educationally deprived children" (Section 205 (a)(5), P.L. 89-10). Over the next several years local evaluation requirements were added to other OE program authorities, and by 1970 several OE bureaus had designated evaluation coordinators whose role was to oversee local evaluation efforts and occasionally to conduct small studies at the national level, usually relying on OE general administrative funds (under the "Salaries and Expenses" account in the annual appropriation) for financial support of any contracted projects.

The fiscal 1970 appropriation for OE contained for the first time, however, a $9.5 million line item for OE evaluation and planning activities. Also in that year John W. Evans was named to head the first OE-wide evaluation office to oversee the expenditure of those funds. To administer a centralized evaluation and planning function, Evans assembled an evaluation staff, composed largely of the evaluation coordinators who had

been working at the bureau level, and also consolidated the various other pockets of federal funds that had until then been sources of bureau-level evaluation support. After that beginning, the activities of the evaluation office grew steadily for the next several years.

With the legislative creation of NIE in 1972, the organizational structure for OE studies of eduation programs was altered somewhat. With a few exceptions, those OE functions that were primarily research oriented were transferred to the new agency. Notable exceptions were the research activities carried out as an adjunct to the OE program for the education of handicapped children.[1] The director of the program argued that the research activities for the education of the handicapped were so closely related to state and local program support activities that handicapped research should not be moved to NIE. The OE handicapped office was successful in this argument and thus paved the way for the 1975 legislative directive in the Education of All Handicapped Children Act (P.L. 94-142) that the major national evaluation activities required in the Act were to be administered by the OE Bureau of Education of the Handicapped (BEH) and not in the central OE evaluation office.

The move towards decentralization of evaluation functions was underscored by language specifying that the new national center for research in vocational education was to be lodged in OE. This action had implications for OE evaluations because the research center was given specific responsibilities for developing evaluation methods and assisting state and local agencies in implementing program evaluations. In the trend towards decentralization of evaluation activities, it was equally important that Congress specified in the vocational education statute (Section 160 (a)(1)) that "the administration of all the programs administered by this Act" was to be the responsibility of the Bureau of Occupational and Adult Education (BOAE). Thus, the management of the national vocational research center and its mandated evaluation activities were explicitly assigned to the OE operating bureau, not to the central evaluation office or to NIE.

The most recent step in this trend has been the shift of responsibility for evaluation of the Follow Through program. As a result of a short-term "exploratory" evaluation of the program, the OE Commissioner in 1979 decided to move Follow Through evaluation activities from

the central evaluation office to the Follow Through
program office. This move was sought by Follow Through
program staff for the stated purpose of making the Follow
Through studies more relevant to program operations.
Undoubtedly, another factor was displeasure of the staff
with a recent large evaluation of the impact of Follow
Through services on student development, reflecting a
frequent pattern of program office/evaluation office
tension (noted in the final section of this paper).

    In addition to the handicapped, vocational, and Follow
Through evaluation activities, OE's evaluation function
had been decentralized in several other ways, even before
the new ED was created. The evaluation office, for
example, has invited the participation of program
managers in all major decisions affecting evaluations in
their respective program areas. The evaluation planning
process, described in the following section, relies
heavily on the judgments and recommendations of program
managers. The importance of this consultation is in some
senses highlighted by the increase in statutory
set-asides of annual program appropriations for national
evaluations. The Emergency School Aid Act of 1972 (P.L.
92-318) specified a set-aside of up to 1 percent of
annual appropriations for national program evaluations.
Two years later, the 1974 reauthorization of ESEA Title I
authorized up to one-half of 1 percent of annual Title I
appropriations for program evaluation and studies. In a
slightly different pattern, the 1974 reauthorization of
ESEA Title VII established a new "Part C - Supportive
Services and Activities" to be administered by the HEW
Assistant Secretary for Education. The 1978 amendments
to Part C authorized studies that are clearly evaluative
in nature, including studies of Title VII effects on
students with language proficiencies other than English
and of methods for identifying students to be served by
Title VII projects. Because the statute assigned
administrative authority for Part C to the HEW Assistant
Secretary for Education, the OE evaluation office was
only one of four offices that has in the past several
years reviewed plans for bilingual activities; the other
offices have been the OE Office of Bilingual Education,
NIE (since it is given specific statutory
responsibilities under Part C), and NCES (since it
conducts statistical studies supporting Title VII).
Under the new Department of Education, the Part C
coordinating function is being carried out by the Office
of Bilingual Education and Minority Language Affairs.

EVALUATION PLANNING

One of the most difficult problems affecting program
evaluation efforts in the Education Division and in ED
has been determining the best way to identify program
evaluation needs.[2] The problem is largely one of
organization. Program managers need to be consulted
regarding any studies to be done in their respective
program areas, and in fact the ED evaluation office has
been consistently careful to ask for the suggestions of
program managers. Program managers and evaluation
managers often disagree, however, with regard to
evaluation priorities for a given program. Program
managers are more likely to ask for evaluation studies
that will help them improve existing management tools or
will enlarge their information about their program
operations; evaluators tend to be more concerned with
whether or not a program is effectively meeting a
longer-range objective, such as the improvement of
academic achievement (or college enrollment rates or
English proficiency) for a defined group of students.
Program managers may not place a high priority on
evaluations of program effectiveness because they believe
that first-order questions (e.g., "Are the intended
children receiving the intended program service?") should,
be answered first or because they fear the consequences
of unfavorable answers to program effectiveness
questions. In addition to this disagreement over the
purposes of evaluations, another organizational problem
is that senior-level program managers often simply are
not willing to take the time to consider evaluation
priorities at the time that decisions must be made.

The OE, now ED, evaluation office has addressed this
need for program consultation by seeking formal
suggestions for evaluations from program managers once a
year. Through 1978 the strategy was to issue an annual
request for project recommendations from program managers
and then to use those recommendations as one factor in
developing a list of projects to be undertaken in the
following year. This list was then submitted to the HEW
Assistant Secretary for Planning and Evaluation (ASPE)
for final approval. The amount of scrutiny by ASPE
varied from year to year; generally only the central
evaluation unit's plans were subjected to critical review
even though other units, such as NIE and NCES, also
submitted their plans.

In 1979 a new procedure was initiated that imposed
greater top-level control over evaluation planning and
was intended to make plans more responsive to concerns of
Congress and senior HEW and ED policy makers. The main
foci of this attention were the proposals of the OE, and
then ED, central evaluation office, but the senior-level
review group convened for the purpose also reviewed
fiscal 1979 evaluation plans prepared by BEH, BOAE, and
the Bureau of Student Financial Assistance (BSFA). The
plans of the central evaluation office, which received by
far the major portion of the group's time and concern,
were criticized and modified by the group primarily with
regard to the proposed timing of studies and their
expected cost; in a few instances plans for impact
studies were delayed by the group and program needs
assessment projects were suggested to precede impact
studies. The group's primary objective with regard to
timing was that new evaluation studies should be
scheduled to provide useful program data in time to make
substantive contributions to legislative debates on
program reauthorization. Cost considerations entered the
decisions to reduce the scope of tasks proposed in
certain studies and to eliminate some tasks from other
studies. Preliminary studies of program need were
recommended in instances in which policy questions
existed about the national need for the type of services
to be provided by the program under review. The new
review procedure was also used for 1980. The resulting
evaluation plan marked the first time that a
comprehensive OE-wide plan had been assembled.

An example of the new review procedure in action was
the group's decision on the proposal of the evaluation
office to examine the effectiveness of the developing
institutions program, Title III of the Higher Education
Act (P.L. 92-318, amended by P.L. 94-482). In that
action the group decided that it was premature to
consider the effectiveness of the program in improving
the financial and educational viability of the
institutions being funded. The group decided that a
necessary first step was to identify a set of reliable
indicators to apply to the financial status of a college
or university in order to determine the financial
strength or weakness of the institution under review. It
was also determined that an "exploratory evaluation" of
the developing institutions program should be
conducted.[3] The purpose of the exploratory study would
be to identify practical, usable measures of successful

project implementation. If such measures could be
identified, it would then be reasonable to go forward
with a larger-scale study, which would--among other
things--actually measure whether or not the developing
institutions program was being fully implemented by
institutions receiving awards under the program.

Under ED Secretary Shirley Hufstedler, the
organizational setting for program evaluation reflected
the increased emphasis on linkages between evaluation and
program improvement. The central evaluation office in ED
reported organizationally to the Deputy Assistant
Secretary for Evaluation and Program Management, who in
turn reported to the Assistant Secretary for Management.
The Program Evaluation Office was organizationally
coequal to the Management Evaluation Office, which was
assigned responsibility for management evaluation,
management quality assurance, program assessment, and
organizational development. In his statement before the
Senate Human Resources Committee prior to confirmation,
John Gabusi, Secretary Hufstedler's Assistant Secretary
for Management, expressed his intent to improve the use
and usefulness of ED evaluations for purposes of
management improvement in ED programs, decisions on
program budgets, and fulfilling information needs of
Congress prior to legislative reviews.

Gabusi's statements and the structure within which the
program evaluation function was organizationally housed
at that time reflect to a considerable extent the
priorities expressed in Circular No. A-11. issued by the
U.S. Office of Management and Budget in March 1979.
Entitled "Management Improvement a d the Use of
Evaluation in the Executive Branch," this directive to
federal agencies construes program evaluation as a
component of federal management improvement. As stated
in the circular, "agency evaluation systems . . . should
focus on program operations and results. They should
include procedures to assure that evaluation efforts
result in specific management improvements that can be
validated" (page 2). The organizational structure under
Secretary Hufstedler reflected these priorities and may
have indicated the direction of upcoming ED evaluation
activity. No information is available at this writing,
however, on the program evaluation plans of Terrel Bell,
Hufstedler's successor as Secretary of Education.

The evaluation of federal education programs has
undergone considerable change in the 10 years in which it
has been a major federal activity. These changes have

included increases in legislative priority on evaluations at federal, state, and local levels. Organisationally, we have seen the federal evaluation function centralized into a single agency-wide unit and then gradually decentralized to some degree. The creation of the Education Department may be quickening the pace of change that characterizes this process. Given these circumstances, it is essential that the direction and character of federal education evaluations be informed by expert; dispassionate analysis of possible methods for increasing the utility of federal evaluations as a tool for improving education.

NOTES

1   A second important exception was the policy research activities carried out by the Education Policy Research Centers. At the recommendation of Evans in 1972,.those ceners (three in number at that time) were moved from the OE evaluation office·to the newly created Office of the Assistant Secretary for Education in order to support that office's activities in education policy development.
2   A similarly difficult issue has been the utilization. of evaluation findings. This issue is addressed in Boruch and Cordray (1980) and in the report of the Committee.
3   Such studies were also undertaken in a number of other program areas at the instigation of Joseph Wholey, ASPE Deputy Assistant Secretary, who had developed the notion of exploring the "evaluability" of a program before full evaluations were done.

REFERENCES

Boruch, R.F., and Cordray, D.S.   (1980) An Appraisal of Educational Program Evaluations: Federal, State, and Local Agencies. Prepared for the U.S. Department of Education. Northwestern University.
David, J.L.   (1980) Federally Mandated Evaluation Policy: The Evolution of the Title I Evaluation Program. Unpublished paper submitted to Education Planning Division, U.S. Department of Health, Education, and Welfare, Washington, D.C.

Frankel, J., Sharp, L., Biderman, A. (1979) Performers
of Research and Research-Related Activities in the
Field of Education. Washington, D.C.: Bureau of
Social Science Research, Inc.

Gordon, R., Leeke, B., Shook, D., eds. (1979) Projects
in Progress - FY 1978. Prepared for Coordinating
Committee on Research in Vocational Education and
sponsored by OE Bureau of Occupational and Adult
Education. Columbus, Ohio: National Center for
Research in Vocational Education, Ohio State
University.

Lyon, C.D., Doscher, L., McGranahan, P., and Williams,
R. (1978) Evaluation and School Districts. Los
Angeles, Calif.: Center for the Study of Evaluation,
University of California.

Office of Management and Budget (1979) Management
improvement and the use of evaluation in the executive
branch. Circular No. A-117. Office of Management and
Budget, Washington, D.C.

Office of Planning, Budgeting, and Evaluation, OE.
(1978) Executive Summary: Student Economic
Background, Achievement Status, and Selection for
Compensatory Services. Washington, D.C.: U.S.
Department of Health, Education, and Welfare.

Orland, M. (1980) Evaluating ESEA Title I: Local
Responses to Federal Expectations. Unpublished paper
prepared for the 1980 meeting of the American
Educational Research Association.

U.S. Congress (1978a) A Report on the Education
Amendments of 1978, H.R.15. House report 95-1137.
Committee on Education and Labor. 95th Congress, 2nd
Session. Washington, D.C.: U.S. Government Printing
Office.

U.S. Congress (1978b) Education Amendments of 1978:
Report to Accompany S.1753. Senate report 95-856.
Committee on Human Resources. 95th Congress, 2nd
Session. Washington, D.C.: U.S. Government Printing
Office.

APPENDIX
B
Performers of Federally Funded Evaluation Studies
Laure M. Sharp

INTRODUCTION AND DATA BASE

The evaluation of federally funded social initiatives in
education--as in health services, crime control, or
housing programs--is seldom carried out by federal
agencies. The bulk of evaluation performers are private
research firms, academic bureaus, and state and local
agencies, which receive federal funds to conduct
evaluations commissioned by congressional mandate or by
executive policy makers or to carry out evaluations on
their own initiative with federal support. Although much
has been written on evaluation methodology and quality,
on one hand, and on the uses and abuses of the grant and
contract system under which federal funds are channeled
to outside performers, on the other, there is no single
useful data base that provides figures on federal funds
spent in a given fiscal year on evaluation activities,
the portion of such funds allocated to outside
contractors or grantees, and the identification of
contract and grant recipients.

Evaluations in the field of education represent a
large share of all federally funded evaluation
activities, probably on the order of one-fifth or
one-fourth of those activities.[1] More specific
information exists with respect to the performers of

217

educational evaluations than for all other evaluations
funded by the federal government; even in education,
however, information is not nearly as extensive and
reliable as one would need for a comprehensive
assessment. The procedure of piecing together relevant
information from various sources is subject to a high
degree of imprecision for several reasons:

• There is no commonly accepted definition of
evaluation activity. In particular, the boundaries
between evaluation and research are far from clear-out,
as discussed by Reisner in Appendix A and by Abramson
(1978) in his work on federal funding of social reserch
and related activities. Evaluation performers themselves
are even more inconsistent with respect to these
boundaries.
• The data that are available seldom refer to the
identical time span. Yet the volume and nature of
federally funded evaluation activities in education have
varied considerably over the time period (1974-79)
considered in this paper.
• While evaluation studies commissioned by federal
agencies have been increasingly funded in the form of
contracts awarded through the competitive procurement
process, work in the evaluation area is also awarded in
the form of grants and "sole-source" awards. In
addition, existing contracts and grants are often
extended and modified, frequently with the addition of
new funds. Information about these types of funding
activities is difficult to locate.
• The prevailing revenue-sharing model under which
large funds are allocated to state and local
jurisdictions on a discretionary basis makes it almost
impossible to estimate the level of evaluation activities
carried out by these jurisdictions. In particular,
systematic documentation is lacking about the extent to
which such activities are performed by staffs of state
and local education agencies or under grant and contract
arrangements by outside organizations. While there is
some discussion in this paper of the evaluation
activities of state and local education agencies, data
presented for those sectors should be viewed as
especially rough estimates.
• While many contracts or grants may be awarded for
the exclusive purpose of conducting an evaluation, there
are probably many more instances where evaluation is
merely one component of a project. This is especially
true of social experiments and demonstration programs.

Most of the data in this paper were obtained through a
survey of performers of research and research-related
activities in education, the American registry of
research and related organizations in education (ARROE).
The ARROE project was conducted from 1976 to 1979 by the
Bureau of Social Science Research under contract to the
National Institute of Education. To create a listing of
potential performing organizations, a variety of sources
was used, including rosters of state departments of
education, intermediate education agencies, local school
systems, federal grantees and contractors, and authors of
articles in 82 pertinent journals. The ARROE project
initially identified more than 6,300 organizations that
might meet the criteria for inclusion in the survey, and
a questionnaire was mailed to each organization.
Organizations that had been active performers during
their last completed fiscal year and were distinct
organizational entities were considered eligible for the
survey and were asked to complete the entire
questionnaire. Organizations that failed to respond were
contacted by telephone, and, if eligible, were asked a
number of key questions. Of the 6,346 organizations on
the original mailing list, 81 percent were contacted and
their eligibility established. Of the 5,208
organizations with whom contact was made, data from just
about half (2,434) were included in the data analysis;
most of the others were ineligible, frequently because
they had not carried out educational RDD&E during their
most recent fiscal year. (The derivation of the ARROE
data base is sketched out in Table B-1.) Slightly less
than half of the reporting units had returned the
detailed mail questionnaires, while slightly more than
half of the units were asked the abbreviated set of
questions in a telephone interview. Thus, the ARROE
survey yielded two data sets: a basic set for all
organizations for whom some data was obtained (N = 2,434)
and a more detailed set (N = 1,071) limited to those
organizations that completed mail questionnaires.[2] The
2,434 performing organizations covered by the survey were
located in 1,530 separate institutions (see Table B-2).[3]

While evaluation was one of the activity areas covered
by the ARROE survey, it was not its primary focus. The
ARROE staff--in consultation with an advisory committee
on which the principal types of performers were
represented--came to the conclusion that in fact most
organizations that perform research and research-related
activities would find it difficult to differentiate

TABLE B-1   The ARROW Data Base

	Organizations in Public Education Agencies	Organizations in Colleges or Universities	Other Organizations[d]	Total
Mail out (N)	1,300	1,655	1,391	6,346
Percentage of returned usable mail questionnaires	35	22	22	24
Percent eligible for ARROW	23	16	14	17
Percent not eligible for ARROW	12	6	8	7
Percentage with whom telephone interviews were completed	50	35	41	39
Percent eligible for ARROW	30	19	20	21
Percent not eligible for ARROW	20	16	21	18
Percentage eliminated after initial contact[b]	3	23	16	18
Percentage of no contact made[c]	12	20	21	19

[a]Includes private profit organizations and organizations in not-for-profit research firms, hospitals, museums, etc.

[b]Organization defunct; organization covered by other reporting unit.

[c]Includes refusals; not reached; in process at time of cutoff.

TABLE B-2 Organizations and Institutions Active in
Educational RDD&E, 1976-77, and Included in ARROE

	Number of Separate Organizations Identified	Number of Institutions in Which These Organizations Were Located	Types of Institutions
Public education agencies	688	631	37 State educa- tion agencies 193 Intermediate service agencies 401 Local educa- tion agencies
Academic	1,268	423	Public and private junior colleges, 4-year colleges, universities, and their divisions; educational R&D centers
All others	478	476	Private nonprofit and for-profit organizations and noninstructional governmental agen- cies; independent education R&D laboratories

between types of functions in funding, expenditures, and
staffing. This was believed to be the case especially
with respect to basic versus applied research, but also
for research versus evaluation and policy studies.
    The definition of evaluation studies also posed a
problem. The ARROE staff and their advisors saw the need
for a fairly restrictive definition, given the propensity
of some respondents, especially those in public education
agencies, to include under the heading of evaluation the
compilation and reporting of periodic or routine
statistics and information. For this reason, ARROE
labelled the relevant category "evaluation and policy
studies," which was defined as: "systematic inquiries
specifically addressed to policymakers and intended to

inform their major policy decisions. Subsumed are
assessments and effects of RDD&E-based programs,
determination of the feasibility of new programs and
projects, and studies focusing on needs, goals, and
priorities of action regarding ongoing or contemplated
activities." Thus, ARROE's definition of evaluation
activities differs to some extent from those used by
other investigators and especially by the Committee.
With respect to the latter, ARROE's definition is both
more restrictive, because it specifies policy makers as
the audience, and broader, because it specifically
includes policy studies.

Using the ARROE definition, several questions about
evaluation activities were included in the mail
questionnaire. Respondents were asked to estimate what
percentage of their education research, development,
dissemination, and evaluation (RDD&E) expenditures were
used primarily for evaluation and policy studies and how
many full-time and part-time professionals spent the
greatest percentage of their working hours performing
evaluation and policy studies. "Project and program
evaluation" was also listed as one of more than 50
problem areas among which respondents could select those
to which their organizational activities were primarily
directed.

The discussion on the following pages is based on
these data and on related analyses of the ARROE data-base
(Frankel 1979, Frankel et al. 1979, Lehming 1979, Sharp
1979, Sharp and Frankel 1979). I believe that this
discussion is helpful in providing a rough picture of the
performer universe and especially of those organizations
that are most active in what is sometimes called the
evaluation industry. It would be foolhardy to claim a
high degree of precision for the numbers presented
here--given such problems as missing data, reluctance on
part of some performers to respond in detail to questions
on financial affairs and on staffing, and possible
respondent misinterpretation or distortion.
Nevertheless, there is enough consistency within the data
set and enough congruence between the ARROE-based
findings and those of other investigators to provide
reasonable confidence about the general trends portrayed
by the data.

ESTIMATE OF FUNDS EXPENDED BY EVALUATION PERFORMERS

On the basis of the ARROE data, I estimate that about
$100 million in federal funds were spent for education
evaluation in 1977 by extramural performers. These
estimates are based on three calculations. First, data
from 80 percent of the 2,434 eligible ARROE respondents
showed aggregate total expenditures for all education
research and research-related activities of $735
million. Adjusting this number for the 20-percent
nonresponse, I estimate total RDD&E expenditures by
educational research performers in 1977 at $900 million.
Second, data from a subset of respondents (864
organizations that completed all relevant items on the
detailed mail questionnaire and reported actual
expenditures of $355 million) showed that approximately
22 percent of all RDD&E expenditures were devoted to
evaluation and policy studies (see Table B-3). Applying
this proportion to the total ARROE population, I estimate
that total expenditures for evaluation and policy studies
in education were approximately $200 million.[4] Third,
about half of all reported RDD&E expenditures in 1977
came from federal sources. This proportion may be a
conservative estimate for evaluation given the
characteristics of the principal performers (which is
discussed below).

Thus, I estimate that in 1977, extramural performers
spent at least $100 million for federally funded
evaluation and policy studies. This figure is
considerably higher than one would derive for 1979 using
Reisner's data in Appendix A, and it is also much higher
than that derived from an available inventory of
competitive contracts awarded by the education agencies
in HEW for fiscal 1977 (Kooi et al. 1978); see Table
B-4. Nevertheless I am reasonably confident that the
figure may be a valid order-of-magnitude estimate for
1977 for several reasons: more funding was available in
1977 than in 1979 (see Table B-4); Reisner's data do not
include expenditures by public education agencies (SEAs
and LEAs), which accounted for a sizable proportion of
all funds expended; Kooi's data do not reflect grants and
sole-source awards, nor do they include continuing work
based on contracts and grants awarded in earlier years,
including supplements made through contract
modifications, while the ARROE study did include funds
for continuations and supplements; the ARROE study also
included performers who received funds from agencies

TABLE B-3 Functional Distribution of Evaluation Expenditures by Sector

Sector	$ Thousands	Research	Develop-ment	Dissemi-nation	Evaluation and Policy Studies	Other	Number of Organizations	Percent
Private								
Profit	31,208	9.3	25.2	10.5	54.8	0.2	22	100
All other[a]	95,277	30.5	22.6	22.7	20.9	3.3	131	100
Total	126,485	25.3	23.2	19.7	29.3	2.6	153	100
Academic	147,086	41.6	24.4	16.4	11.5	6.2	474	100
Public								
Small LEAs	11,433	19.8	29.3	7.2	33.9	9.9	109	100
Large LEAs	20,464	12.3	25.6	7.6	48.6	5.8	34	100
ISAs	12,896	12.1	25.9	31.6	29.2	1.2	55	100
SEAs	35,344	14.2	42.4	15.4	22.5	4.9	36	100
Total	80,137	14.1	32.6	16.1	31.9	5.3	234	100
TOTAL	354,490[b]	29.5	25.8	17.7	22.4	4.7	864	100

[a]Includes primarily private nonprofit organizations, including independent nonprofit R&D organizations and public organizations (e.g., state and local agencies outside the field of education, such as hospitals or health agencies), as well as those organizations whose profit or nonprofit status could not be determined because they did not supply the information.
[b]Includes $782,000 not identified by sector.
SOURCE: ARROE mail questionnaire respondents only.

236

TABLE B-4 Competitive Procurements in 1977 and 1979 for
Evaluation Studies by Sector

	1977[a]	1979[b]
Public agencies	--	45,000
Academic institutions	199,000	38,238+
Private (profit or nonprofit)	5,326,654	2,664,613
TOTAL	$5,525,654	$2,747,851

[a]Data from Kooi et al. (1978).
[b]Preliminary data from Kooi et al. (In press), made available to
the author.

other than HEW (for example, from DOL or NIH) for work
that could be classified as education RDD&E; and
classification differences--in particular the inclusion
of policy studies--may have inflated the evaluation
estimates for ARROE.

SELECTED CHARACTERISTICS OF PERFORMING ORGANIZATIONS

Who were the performers of evaluation work in 1977 and
how were federal funds for evaluation distributed among
various sectors of the performer community? For analytic
purposes, ARROE classified the performer community into
three major segments: the public education sector, which
included state education agencies (SEAs), intermediate
service agencies (ISAs), and local education agencies
whose enrollment was 10,000 or more, which in turn were
subdivided into large LEAs (with enrollments of 50,000 or
more) and small LEAs (with enrollments of 10,000-49,000);
the academic sector, which included public and private
two-year and four-year colleges, universities, and their
subdivisions, such as R&D centers, specialized
institutes, and survey units; and a residual sector,
which was largely composed of profit and not-for-profit
research and development organizations and educational
laboratories, but also included hospitals, publishers,
foundations, associations, and noneducational agencies of
state and local governments, such as health and manpower
agencies.
   As shown in Table B-5, academic organizations
represent the largest single group of performers of
educational research and related activities, followed by

TABLE B-5 Distribution by Sector of All RDD&E Performer and of Evaluation Performers

		All RDD&E		Evaluation	
Sector	(N)	$ Thousands	Percent	$ Thousands	Percen
Private					
Profit	(22)	31,208	8.8	17,094	21.5
Other	(131)	95,277	27.1	20,151	25.3
Total	(153)	126,485	35.9	37,245	46.8
Academic	(474)	147,086	41.4	16,911	21.2
Public					
LEA--small	(109)	11,433	3.2	3,870	4.8
LEA--large	(34)	20,464	5.8	9,953	12.5
ISA	(55)	12,896	3.6	3,778	4.7
SEA	(36)	35,344	9.8	7,873	9.9
Total	(234)	80,137	22.4	25,474	32.0
TOTAL	(864)	354,490[a]	100.0	79,645[b]	100.0

[a]Includes $782,000 not identified by sector.
[b]Includes $15,000 not identified by sector.
SOURCE: ARROE mail respondents only.

those in the private sector. Public education agencies accounted for less than one-fourth of all RDD&E expenditures.[5] With respect to evaluation, however, the picture is very different. Organizations in the private sector were in first place, followed by public education agencies, and academic performers had the smallest share. Furthermore, as shown in Tables B-6 and B-7, only in two types of organizations--private for-profit and local school systems--is there a concentration of organizations that spent more than $100,000 on evaluation in 1977 or devoted most of their resources (50 percent or more) to evaluation activities. The data clearly suggest that evaluation is a marginal activity for most academic performers, while it plays a major role in sustaining most for-profit organizations. However, given the actual numbers of performers involved, one should not conclude that most large evaluation dollars were spent by private for-profit organizations in 1977: 5 for-profit organizations spent in excess of $500,000 for evaluation compared with 12 not-for-profit

TABLE B-6   Level of Expenditures for Evaluation by Reporting Organizations (percent of organizations)

| Level of Expenditure (dollars) | All Organizations | Type of Organization | | | | | | |
| | | Private | | Public | | | | |
		Profit	All Others	Academic	SEA	ISA	LEA-- Small	LEA-- Large
0	30.6	17.4	29.0	40.2	10.8	23.7	11.7	5.9
$1-24,999	31.6	8.7	27.5	32.2	18.9	45.8	41.4	11.8
$25,000-99,999	21.0	21.7	15.3	18.2	27.0	20.4	37.8	20.6
$100,000-500,000	13.7	30.4	21.4	8.6	35.1	8.5	9.0	47.1
Over $500,000	3.1	21.7	6.8	0.8	8.1	1.7	--	14.7
TOTAL (percent)	100.0	99.9	100.0	100.0	99.9	100.1	99.9	100.1
Number of cases	873	23	131	478	37	59	111	34

SOURCE: ARROE mail respondents only.

TABLE B-7 Percentage of Organizations' Total Expenditures
Devoted to Evaluation

Sector	O	1-24	25-50	50+
Private				
For profit	19.2	15.4	26.9	38.5
All other	23.7	43.9	18.0	14.4
Academic	34.9	41.4	11.6	12.0
Public				
SEA	14.0	51.2	25.6	9.3
ISA	18.6	52.5	15.3	13.6
LEA--small	8.8	24.6	26.3	40.4
LEA--large	5.4	13.5	37.8	43.2

SOURCE: ARROE mail respondents only.

organizations and 4 academic organizations that spent
that amount.

There are sharp differences among organizations in the
various sectors of the performing universe. The balance
of this section examines separately some salient features
of evaluation performers in each of the three sectors.

For-Profit and Not-for-Profit Organizations in the Private Sector[6]

What is sometimes referred to as the evaluation industry
is a group of organizations--some profit, some
not-for-profit, some large, others quite modest--that are
at present the most frequent performers of federally
funded evaluations in the field of education. With the
emergence and the predominance of the competitive
procurement system and the funding of evaluations under
contracts rather than grants, organizations of this type
are apparently best able to mount the prodigious proposal
writing efforts required for participation in the system
and to muster and manage the resources necessary to carry
out large-scale evaluation projects, often under severe
time constraints.

Obviously, the ARROE data collection effort, since it
was not targeted to performers of federally funded
evaluation but sought instead to capture the universe of
organizations that contributed to research, development,
and evaluation in education in 1977, failed to isolate

the full set of organizations that are of interest for an
assessment of federally funded evaluation performers.
Nevertheless, some of the findings are instructive: 211
of the 478 organizations in the residual sector (i.e.,
affiliated neither with academic institutions nor with
public education agencies) were classified as R&D
organizations and thus constitute the universe of
organizations potentially involved in the "evaluation
industry" (see Table B-8). Most of these 211
organizations spent less than $1 million on all research
and research-related activities in 1977, regardless of
source of funding. The 77 organizations that spent $1
million or more in 1977 include the federally funded
educational laboratories (a group of not-for-profit
institutions started with federal funding but now partly
dependent on grant and contract work) and a number of
not-for-profit groups primarily oriented to the field of
education or educational administration. The ARROE data
are incomplete (about one-third of the respondents did
not wish to disclose the information or have their names
associated with the information if they did disclose it)
but no more than 15 organizations were identified that
are members of the "industry" as popularly conceived
(System Development Corporation, Abt Associates, American
Institutes for Research, Educational Testing Service
(ETS), etc.). Only three such organizations are among
the 10 private-sector organizations that reported
expenditures of more than $5 million for all education
RDD&E; the other 7 organizations were educational
laboratories, not-for-profit education centers, and
hospitals, presumably engaged in research centered on the
education of medical personnel.[7]

More than other organizations, those in the private
sector and especially the major performers depend heavily
on federal funding for their activities. According to
the ARROE study, 62 percent of the funding for the
private sector came from federal sources compared with 48
percent for the academic sector. Academic institutions
rely to a greater extent on state and local government
funding: 19 percent of education RDD&E work in the
academic sector was funded from state and local sources,
but only 10 percent of the work in the private sector.
Large private-sector organizations and organizations that
specialize in education RDD&E in particular have few
other sources of funding: half of the organizations that
spent more than $1 million in 1975 for education RDD&E
received at least 75 percent of their funds from the

TABLE B-8   Types of Organizations in Private Sector[a]

	All Organizations		Organizations Spending $1 Million or More	
	N	Percent	N	Percent.
Education RDD&E	155	35	26	47
Other RDD&E	56	13	9	16
Non-RDD&E	213	48	18	33
Health care	50	11	3	5
Associations, labor unions	35	8	3	5
Private schools	24	5	--	--
Social science	17	4	--	--
Child care	16	4	--	--
All others	71	16	12[b]	22
Government agencies	23	5	2	4
TOTAL	447[c]	100	55[d]	100

[a]Includes government agencies other than public education agencies.
[b]Publishing, Broadcasting--2. Management Consulting--2. Information Services--2. Other--6.
[c]Information not available for 31 cases.
[d]Information not available for 3 cases.
SOURCE: ARROE mail and telephone respondents.

federal government, and one-fourth of them received at least 90 percent from the federal government.

The ARROE data show that large performers (expenditures of $1 million or more) account for the bulk of all expenditures in education RDD&E in the private sector; while they are 18 percent of all organizations listed in ARROE, they accounted for 77 percent of all reported expenditures. For the subset of organization for which there are more detailed data, the picture was similar; furthermore, expenditures for evalution are even more heavily concentrated among major performers than are expenditures for all RDD&E (see Table B-9). But these performers do not fit the image of an industry whose only activity and source of revenue is the performance of evaluations in the field of education: federally funded evaluation work is concentrated in large organizations with diversified activities that encompass various topical areas (for example, the Rand Corporation, Abt Associates, and Applied Management Sciences) or several different research functions or activities in education (for example, ETS).

TABLE B-9 Distribution of Total Expenditures and
Evaluation Expenditures in Private and Academic Sectors
by Major and Minor Performers[a]

	Total RDD&E Expenditures in 1977		Expenditures for Evaluation in 1977	
	Percent	Number	Percent	Number
Private sector				
All organizations	100.0	354	100.0	153
Major organizations	79.6	58	82.7	32
All other organizations	20.4	296	17.3	121
Academic sector				
All organizations	100.0	943	100.0	474
Major organizations	50.1	92	46.1	39
Minor organizations	49.9	851	53.9	435

[a]Major performers are those who spent more than $1 million for
all RDD&E activities in 1977; minor performers are all others.
NOTE: "Total RDD&E Expenditures" column is based on both mail
and telephone respondents. "Expenditures for Evaluation" column
is based on mail respondents only. All cases with missing data
were excluded.

This is not to say that one or another organization
may not have come into existence for the purpose of only
such activities—or even for the purpose of performing a
single contract with a given agency, a pr
highlighted in a recent GAO report,[8] especially with
respect to former employees (U.S. General Accounting
Office 1980). Small performers do carry out a fair,
amount of educational research and research-related work,
and some may fit the image of the "beltway bandits" so
prominently mentioned in all the periodic exposes of the
research and contract world. It is also possible that
such respondents were especially unlikely to return the
ARROE questionnaire and were interviewed by telephone and
so were underrepresented in the group from whom detailed
information was obtained. However, the evidence
indicates that the bulk of evaluation work is done by a
relatively small number of well-established and fairly
large organizations. This hypothesized distribution of
activities across types of organizations is confirmed by
an (incomplete) inventory of competitive evaluation
contract awards made in 1977 and 1979 (Kooi et al. 1978,

in press). An earlier study by Biderman and Sharp (1972)
led to similar conclusions: while it identified a large
number of active organizations in the competitive
procurement process, it found that awards for the
unrestricted, open procurements most often went to very
active bidders, usually large organizations. Since 1972,
with increasing emphasis on open competitions, this trend
has no doubt accelerated.

As is shown in the next section of this paper, the
major performers of evaluations have large professional
staffs drawn from a wide range of disciplinary
backgrounds. Less is known about the smaller
organizations that perform the balance of federally
funded evaluations; their activities and staffing
patterns are largely undocumented since they have not
become part of the professional and disciplinary networks
in which the large organizations participate.

Evaluation in Academic Institutions

As was shown in Table B-2, evaluation clearly represented
a smaller share of total RDD&E activities for academic
organizations than for other performers. Furthermore,
despite the fact that academic organizations are the
largest performers of all education RDD&E, the dollar
amounts involved in evaluation work were relatively
small. It is not possible to ascertain from the ARROE
data to what extent academic evaluation expenditures were
funded with federal dollars obtained directly through a
grant or contract from one of the education agencies in
HEW or with federal dollars that had gone to a state or
local agency that in turn contracted the evaluation to a
college or university.

When social-science-based evaluation was first used to
assess social programs, academic institutions were
frequent performers of major evaluations, usually under
grant or sole-source contract arrangements. The reasons
for a gradual shift from grants to contracts and from
academic to other types of research performers have been
amply discussed in a number of publications (see, e.g.,
Williams 1972), most recently by Levitan and Wurzburg
(1979), who claim that by 1974 HEW had ruled out further
support of evaluations under grants and that sole-source
contracting became increasingly difficult. They report
that by 1979 officials estimated that less than 10
percent of HEW evaluation funds were awarded

244

noncompetitively. Whether the decline in federally
funded evaluation activities on the part of academic
units is due to their decision not to participate in
competitive procurements or to lack of success when they
do so cannot be ascertained from available data. It is
clear that they do not win many competitive awards:
Kooi's inventories of competitive procurements for 1977
and 1979 showed only one study in each of the two years
that could be unequivocally classified as an evaluation
study competitively awarded to an academic institution.

In their study of evaluation performers, Biderman and
Sharp (1972) found that only 11 percent of the 1,324
organizations identified as RFP recipients were
academically affiliated institutions, and the majority of
these had received the RFP at the agency's initiative. A
total of 225 bids were filed for 36 procurements; only 17
of them were submitted by academically affiliated
organizations; and only one award, not for an evaluation
study, went to an academic organization. These earlier
data suggested that academic organizations did not
participate very actively in the federally organized
competitive procurement system at that time, and this may
not have changed a great deal since.

Evaluation in Public Education Agencies

Federal dollars are spent by state and local public
education agencies primarily to perform evaluations that
are mandated in conjunction with federally funded
education activities. In addition, state or local
agencies may carry out federally funded demonstration or
research projects that have built-in evaluation
components. State or local agencies can also participate
in competitions for evaluation contracts; this is rare,
however, since there are more restricted types of
competitive procurements (for example, for various
demonstration and innovative programs) that are targeted
primarily to public education agencies and hence are
preferred by them.

As shown in Table B-6 above, evaluation occupies a
more prominent place in the activities of local education
agencies than in those of any other sector: more than 40
percent of such agencies included in the ARROE study
indicated that more than half of their research and
research-related activities were devoted to evaluation.
The resources of these education agencies are often

considerable; among the surveyed organizations that
reported spending more than $5 million in 1977, two were
LEAs: Los Angeles and Leon County, Florida. However,
many of the evaluation activities undertaken by such
agencies tend to rely heavily on student tests, so that
the boundaries between "testing" and "evaluation" are
often hard to draw. It may be for this reason, or
perhaps because LEAs do not always identify sources of
evaluation funding accurately, that LEAs appear to be
somewhat less dependent on federal funds than are other
public agencies to carry out their evaluation activities
(see Table B-10).

Evaluation--at least as defined for the ARROE
study--plays a lesser part for state agencies than it
does at the local level, but (as shown in Table B-3
above) the actual amounts involved are larger because of
the higher expenditure levels in these agencies.
Relatively few state and intermediate service agencies
spent more than 25 percent of their RDD&E resources on
evaluation.

According to the National Science Foundation (NSF)
(1980), local personnel generally tend to perform most

TABLE B-10   Percent of All Organizations Reporting That
Half or More of Their Funds Came From Federal Sources
in 1977

	Organization for Which Evaluation was a Major Activity		Organization for Which Evaluation was Not a Major Activity	
	Number	Percent	Number	Percent
Public				
SEA	26	73.1	18	50.0
ISA	37	35.1	23	30.4
LEA--large	33	24.2	5	60.0
LEA--small	84	11.9	28	35.7
Academic	241	36.9	290	39.3
Private				
Major	16	68.6	8	62.5
All other	70	52.9	74	62.2

NOTE: Organization could check more than one "major activity"
area.
SOURCE: ARROE mail questionnaire respondents.

246

research and related activities in-house, although the
portion performed extramurally has increased in recent
years, from 20 percent in 1966 to close to 40 percent in
1977. Of that 40 percent, private firms performed 17
percent; not-for-profit firms, 13 percent; and
universities and colleges, about 10 percent. The extent
to which this pattern holds for education as compared
with energy, environment, health, etc. cannot be
ascertained from the NSF data. However, information from
a recent survey of school districts (Lyon 1978) indicates
that on the average only 6 percent of the budget of a
district's evaluation units was spent on outside
consultants, although there was considerable variation
from district to district. State agencies, too, appear
to perform most work in-house: one recent study reports
that 73.3 percent of all research and research-related
activities are conducted by agency staffs (Mathis and
Walling 1979).

PERSONNEL

The organizations included in ARROE employed
approximately 22,200 full-time and 12,000 part-time
professionals in 1977. The distribution of personnel
matches the distribution of funds, although in the
aggregate, academic institutions allocate more persons
per dollar than organizations in the other sectors (see
Table B-11). Staff qualifications vary by sector, with
those in academic organizations most likely to hold a

TABLE B-11   Staffing and Funding Allocation for Education
RDD&E, by Sector, 1977 (in percentages)

Sector	Full-Time Professionals	Part-Time Professionals	Funding
Private	27	16	33
Academic	58	76	51
Public	15	7	16
TOTAL (percent)	100	100	100
Number	22,286	12,024	$735 million[a]

[a]Based on reports from 80 percent of respondents.
SOURCE: ARROE mail and telephone respondents.

doctorate degree; more surprising, private-sector
organizations are more likely to employ people from a
wider spectrum of academic disciplines (see Table B-12).

As was noted above, most organizations do not
specialize in evaluation, and therefore staff is likely
to be used interchangeably between evaluation and
research. Insofar as the ARROE data allow
differentiation, however, the following characteristics
apply to those staff who actually worked on evaluation
studies in 1977. First, the percentage of total staff
allocated to all evaluation was slightly lower than the
percentage of expenditures: 22 percent of funds and 17
percent of personnel were devoted to evaluation and
policy studies. This is not unexpected since the
staff/dollar ratio for all RDD&E is highest in the
academic sector and lowest in the private sector (see
Table B-11) and the private sector is the most frequent
performer of evaluations. In the absence of data, one
can only speculate about the reasons for the difference
in staff/dollar ratio. It may be due to the greater

TABLE B-12 Selected Characteristics of Full-Time Staff,
by Sector, 1977

	Public	Academic	Private
Percent of full-time staff with doctorates	28	67	31
Percent with major field of expertise in:			
Education	65	58	41
Psychology	9	10	16
Other social science	3	9	12
Humanities	2	2	5
Physical and biological sciences	1	7	2
Mathematics, statistics	7	2	5
Business economics, accounting, public administration	3	2	5
Communications, library science	3	3	7
Operation research, systems analysis	4	1	4
Other	3	6	4

SOURCE: ARROE mail respondents only; response rate to this question was 40 percent.

dependence of private performers on federal funding in
comparison with public and academic institutions that may
be able to cover overhead or some personnel costs from
regular budgets. The availability of low-cost labor
(graduate students and post-doctoral fellows) on many
campuses may also be reflected in these figures; the data
in Table B-11 suggest that academic institutions are able
to take advantage of the availability of faculty or
students for part-time employment. However, the
difference in staff/dollar ratios may also be due to the
fact that private contractors and grantees spend higher
proportions of their funds on nonpersonnel items such as
computer work, which is often available at relatively low
cost in university settings. Another factor may be high
overhead costs in the private sector due, in part, to
proposal writing or marketing costs that are especially
high in that sector.

Second, there are also some noteworthy differences
with respect to staff training. Table B-13, which
presents differences in the presence of doctorate holders
on the staffs of reporting organizations, uses a
different base from most of the other data shown in this
paper. Organizations were categorized according to their
answer to a question about major activity areas, one of
which was program and project evaluation. (Respondents
were free to check as many areas as applied to their
organizations, and most checked more than one.)
Respondents were then classified into evaluators and
nonevaluators based on their answers.[9] Again it is
necessary to bear in mind that not all evaluation
performers are in the "evaluator" category, but only
those who indicated that evaluation was a major
activity. Although in many cases the cell sizes are
quite small, some comparisons can be made: in the
academic sector, the participation in research and
research-related activities of those who have Ph.D.s is
ubiquitous. About three-fourths of all academic units
performing this type of work employ Ph.D.s, whether they
do evaluations or not. In most other types of
organizations, there tends to be at least one person with
a Ph.D. on the staff, but the number of Ph.D.s is greater
if one of the major activities is evaluation work. The
difference is especially striking in public agencies, but
in the private sector, too, evaluation performers almost
always have at least one person with a Ph.D. on the
staff. Only in state agencies does the presence of
evaluation activities not affect staff characteristics:

TABLE B-13   Selected Characteristics of Organizations With and Without Evaluation as a Major Activity

	Private Profit	Private All Other	Academic	Small LEAs	Large LEAs	ISAs	SEAs	Total
Organization for which evaluation is a major activity								
Percent of full-time staff with doctorates								
0	6.7	24.6	8.3	19.7	5.7	20.7	11.5	13.3
1-24	26.7	16.4	5.2	6.6	31.4	24.1	11.5	11.7
25-49	26.7	26.2	10.9	7.9	25.7	24.1	50.0	17.5
50+	40.0	32.8	75.1	65.8	37.1	31.0	26.9	57.5
Number	7	60	199	18	7	15	18	324
Organization for which evaluation is not a major activity								
Percent of full-time staff with doctorates								
0	14.3	31.7	11.1	44.4	28.6	60.0	5.6	19.1
1-24	42.9	10.0	5.0	0.0	14.3	6.7	27.8	8.1
25-49	42.9	21.7	9.0	16.7	57.1	33.4	66.6	13.6
50+	0.0	36.7	74.9	38.9	0.0	0.0	0.0	59.3
Number	15	61	193	76	35	29	26	435

SOURCE: ARROE mail respondents only.

250

there is at least one person with a Ph.D. in most
agencies, regardless of the nature of the work.
For-profit organizations are especially likely to employ
Ph.D.s if they are engaged in evaluation. It should be
noted, however, that the data in this category are from a
small number of organizations.

Equally interesting differences can be observed with
respect to staff specialization, i.e., the presence of
disciplinary specialists on an organization's staff.
Table B-14 shows that organizations for which evaluation
is a major activity tend to have more diversified
staffs. This is especially the case in the private
sector, but holds true in the other sectors as well.

Obviously, staff size, percent of staff with Ph.D.s,
and diversification of disciplines among the staff are
not in themselves a guarantee of efficient or
high-quality performance; in the aggregate, however, they
furnish some indication of the efforts expended by those
who carry out evaluation work within the educational
research community. Generally, the performers of
evaluation activities tend to be organizations with
staffs that are larger, better trained, and more
diversified than the staffs in organizations for which
other types of research and research-related activities
constitute a major activity.[10]

CONCLUSION AND COMMENT

Despite the difficulties of distinguishing between those
who perform evaluations and those who perform other types
of educational research, and between those who are funded
from federal sources and those who are not, some
differences among performers emerge from the ARROE data.
Of greatest interest are differences between academic and
private-sector organizations, since they are the true
outsiders who perform evaluations under federal
auspices. The public agencies are important performers
and their activities are of crucial importance in the
assessment and evaluation of the impact of federal
dollars spent on education, but the mechanisms at the
disposal of the federal government in initiating and
monitoring evaluations in the public sector are very
different from those that apply to contracts and grants
awarded to academic and private organizations.
Furthermore, public evaluation units exist and function
to a large extent in a self-contained universe, while the

TABLE B-14 Percent of Organizations with at Least One
Full-Time Staff Member in Selected Disciplinary Fields
(in percentages)

	Education	Psychology	Math and Statistics	Other
Private profit				
Major evaluation performer	100.0	71.4	35.7	100.0
Other	71.4	57.1	50.0	53.8
Private other				
Major evaluation performer	87.1	42.6	16.2	66.7
Other	65.5	29.3	5.2	66.7
Academic				
Major evaluation performer	78.8	37.6	16.5	52.0
Other	67.7	30.3	12.2	46.5
Small LEA				
Major evaluation performer	85.9	22.5	19.2	33.3
Other	88.9	11.1	11.1	25.7
Large LEA				
Major evaluation performer	85.3	55.9	51.5	16.7
Other	83.3	50.0	23.3	40.6
ISA				
Major evaluation performer	93.1	34.5	31.0	33.3
Other	73.3	13.3	6.7	55.2
SEA				
Major evaluation performer	96.3	22.2	33.3	56.2
Other	94.1	25.0	22.2	70.4
All organizations				
Major evaluation performer	84.5	37.2	73.8	53.2
Other	64.3	29.0	12.4	49.4

SOURCE: ARROE mail respondents only.

two other sectors compete, interact, and cooperate with
respect to much of the evaluation work and related
activities.

It is clear from the ARROE data that academic units
continue to do the bulk of educational research in
general and that large numbers of well-qualified persons
are involved in such activities. Universities have at

their beck and call resources that can be used on a
part-time basis, as the ARROE data clearly show; such
utilization is often economical and advantageous.
Therefore, it seems unfortunate that academic
institutions participate so little in one of the most
important segments of the work being done today in the
field of educational research, namely evaluation. While
private organizations can to some extent duplicate
university staffing arrangements through the use of
consultants, including academic consultants, this often
requires travel, less opportunity for day-by-day
involvement, and higher costs. Such arrangements also
cannot provide the opportunity available at universities
for faculty and graduate students to stay in close touch
with practical problems and federal concerns and for
better articulation between graduate training and
employment requirements.

But it is also worth noting that as a result of the
shift to the private sector, a number of organizations
have emerged that have large, sophisticated,
multidisciplinary staffs that are very knowledgeable
about the major educational issues of the day. Whether
the present federal procurement system leads to the best
possible utilization of these resources is not clear:
earlier research (Biderman and Sharp 1972) and anecdotal
evidence suggest that the timing of requests for
proposals, the imposition of tight deadlines coupled with
time-consuming clearance procedures, and the need to
devote enormous efforts to proposal preparation all
militate against optimal utilization. In any case, the
maintenance of this capability is far from certain, given
the reduction in the volume of federal evaluation
procurements in education and the ability of many of the
private-sector firms to redeploy personnel to areas such
as energy, or transportation, or defense, which may be of
higher priority than education. The loss of these
specialists will be detrimental to the field of
educational research, which has long suffered from a
narrow and parochial perspective.

As the report and other cited sources show, a
convincing case can be made that the current procurement
system is not designed for optimal efficiency.
Increasingly, the choice of grants or contracts as a
means of supporting work is not based on substantive
considerations, and the eligibility criteria (based on
such categorical descriptors as profit or not-for-profit,
minority-owned, etc.) may preclude performance by

well-qualified organizations. The contracting system is
a necessary ingredient of a government process in which a
heavy activity and service load is mandated together with
low federal personnel ceilings (Sharkansky 1980), but it
needs to be made more flexible. The data presented in
this paper suggest that most evaluation work in education
commissioned at the national level is done by performers
who have the experience and resources to perform it well,
despite occasional awards that are open to question (U.S.
General Accounting Office 1980). But the universe of
performers is a relatively narrow one. The
diversification of this universe through greater
participation by university-based research groups, the
preservation of existing proven resources in the private
sector, and improvements in the procurement system should
be of concern to those who seek to increase the quality
and utility of evaluations.

NOTES

1 This estimate is based on Abramson's data (1978),
   which showed for 1977 a total of only $63.6 million
   for all federally funded evaluations. While
   Abramson's definition of evaluation yields a much
   lower estimate of total evaluation activities than is
   generally used by other researchers, this figure can
   be used to gauge the relative shares of expenditures
   by various government agencies. Of the $63.6 million,
   HEW accounted for more than half, with welfare
   agencies accounting for the largest bloc (more than
   $16 million) and education for the second largest
   (close to $14 million).
2 Because of item nonresponse--especially with respect
   to funding questions--the actual numbers of cases
   available for analysis is usually somewhat smaller.
3 Especially in academic institutions, it is not
   uncommon to have several separate, autonomous units
   (for example a school of education, a survey research
   unit, and the department of psychology) performing
   education research and research-related activities.
   Of the 1,268 academic organizations shown in Table
   B-2, the largest number (34 percent) were individual
   departments, followed by divisions or schools (24
   percent) and bureaus and centers (24 percent).
4 The data files were examined for nonresponse bias and
   for mail versus telephone respondent bias, as well as

for error due to missing data (item nonresponse). For
the variables available for this analysis (size of
organization, sector, etc.), there were no obvious
biases, but of course there is the always unanswered
question about characteristics of reluctant
respondents or nonrespondents that demographic
variables do not capture.

5 These data are based on the subset of mail
respondents. Total expenditure data for all ARROE
organizations showed the same ranking and order of
magnitude, but slightly different
percentages--academic 57 percent, private 33 percent,
public 16 percent--suggesting that "active" public
education agencies were more likely to return the mail
questionnaires.

6 As shown in Table B-8, this nomenclature includes a
few government agencies other than public education
agencies.

7 The 10 private-sector organizations that reported
expenditures of more than $5 million (in most cases
for fiscal 1977) are Abt Associates, Inc., Education
Commission of the States, Education Development
Center, Inc., Education Finance Center, Educational
Testing Service, Far West Laboratory for Educational
Research and Development, Montefiore Hospital and
Medical Center, Northwest Regional Education
Laboratory, St. Louis Childrens Hospital, System
Development Corporation.

8 None of the contracts criticized on this basis in the
GAO report were awarded by an education agency.

9 I am indebted to Georgine Pion and Robert Boruch of
Northwestern University for suggesting these
tabulations and making funds available for the
required computer work.

10 But it should be kept in mind that ARROE encompasses a
highly diverse set of organizations, including some
that specialize in development and dissemination, for
which these same characteristics may not be relevant
to work performance.

REFERENCES

Abramson, M.A.   (1978) The Funding of Social Knowledge
Production and Application: A Survey of Federal
Agencies. Washington, D.C.: National Academy of
Sciences.

Biderman, A.D., and Sharp, L.M. (1972) The Competitive
Evaluation Research Industry. Washington, D.C.:
Bureau of Social Science Research, Inc.

Frankel, J. (1979) Education RDD&E in the Public
Education Sector: Highlights of ARROE Findings.
Paper presented at the American Educational Research
Association annual meeting, San Francisco.

Frankel, J., Sharp, L., and Biderman, A. (1979)
Performers of Research and Research-Related Activities
in the Field of Education. Washington, D.C.: Bureau
of Social Science Research, Inc.

Kooi, B., and Washington Representative Services staff
(1978) Annual Analysis of Federal Competitive
Procurements in Management Consulting, FY'77.
Washington, D.C.: Capitol Publications, Inc.

Kooi, B., and Washington Representative Services staff
(in press) Annual Analysis of Federal Competitive
Procurements in Management Consulting, FY'79.
Washington, D.C.: Capitol Publications, Inc.

Lehming, R. (1979) Research Organizations in Education
in Academic Institutions. Paper presented at the
American Educational Research Association annual
meeting, San Francisco.

Levitan, S.A., and Wurzburg, G. (1979) Evaluating
Federal Social Programs--An Uncertain Art. Kalamazoo,
Mich.: W.E. Upjohn Institute for Employment Research.

Lyon, C.D., Doscher, L., McGranahan, P., and Williams,
R. (1978) Evaluation and School Districts. Los
Angeles, Calif.: Center for the Study of Evaluation,
University of California.

Mathis, W., and Walling, P. (1979) A National Survey of
State Educational Agency Research Organizations.
Paper presented at the American Educational Research
Association annual meeting, San Francisco.

National Science Foundation (1980) State and local
government R&D expenditures, FY 1977. Reviews of Data
on Science Resources, NSF 80-302, No. 35(March):14.

Sharkansky, I. (1980) Policy making and service delivery
on the margins of government: The case of
contractors. Public Administration Review
(March/April):116-123.

Sharp, L.M. (1979) Education RDD&E in the Private
Sector. Paper presented at the American Educational
Research Association annual meeting, San Francisco.

Sharp, L.M., and Frankel, J. (1979) Organizations that
perform education R&D: A first look at the universe.
Educational Researcher 8(11):6-11.

U.S. General Accounting Office (1980) Controls Over
    Consulting Service Contracts at Federal Agencies Need
    Tightening. PSAD-80-35. Washington, D.C.: U.S.
    General Accounting Office.
Williams, W. (1972) The capacity of social science
    organizations to perform large-scale evaluative
    research. In P.H. Rossi and W. Williams, eds.,
    Evaluating Social Programs. New York: Seminar Press,
    Inc.

How the Evaluation System Works: The State and Local Levels
Freda M. Holley

Kaleidoscopic is a good term to describe the evaluation
of federal programs at the local and state level. There
is enormous variation both from state to state and from
district to district. Moreover, the practice of
evaluation differs across programs within those states
and within those districts.

This paper attempts to give some flavor of that
variation in such areas as evaluation funding and
budgets, personnel, evaluation activities and practices,
and, finally, in dissemination and utilization. The
paper concludes with some discussion of the implications
of this variation. The reader is cautioned against a
quick assumption that such variation is undesirable: it
may well be that such variation is not helpful to those
making decisions at the federal level, but it must be
remembered that national program success can only be
built block by block at the local level. Considerable
variation may be necessary to foster program
implementation and to respond to differing needs at the
local level. Imagination may be required at the national
level to use such variation creatively to the benefit of
national purposes. It may also be necessary to recognize
that it is pointless to attempt evaluation at the
national level; one evaluation system cannot serve both
the local, state, and national needs. In any case, the

The author, a member of the Committee, is head of the
Office of Research and Evaluation of the Austin
Independent School District, Austin, Texas.

246.

size of the expenditures and the importance of the local
endeavor dictate that the Department of Education should
work to optimize the return from evaluation efforts at
the local and state level, both for local and national
aims.

HOW ARE EVALUATIONS FUNDED?

Our best evidence on the extent of variation in federal
program evaluation at the state education agency (SEA)
and local education agency (LEA) levels is related to
evaluation budgets. Budgets are a major concern in local
and state evaluation efforts, of course, and for this
reason most of the data collection has focused on them.
The most recent data were collected in a survey of state
and large city evaluation units on behalf of a task force
on resource allocation in program evaluation appointed by
Division H (School Evaluation and Program Development) of
the American Educational Research Association (AERA).
This survey (Drezek and Higgins 1980) reported that the
size of LEA budgets for the evaluation of Title I
programs ranged from zero to $935,000 for Title I program
budgets of $104,000 to $52 million. Similarly, the range
of median reported funding expressed as a percent of
program funding across major programs ranged from 7
percent for ESEA Title IVC (innovative practices and
curriculum) to 0.5 percent for P.L. 94-145 (special
education); see Table C-1 for details.

   Doss (1979) surveyed large districts in the Southwest
in order to gather descriptive information about their
Title I evaluation efforts. This survey reveals similar
variation: one program with a $3,563,071 budget had an
evaluation budget of $10,000; another program with a
budget of $2,447,020 had an evaluation budget of $88,036
(see Table C-2). The percentages reported by Doss
closely parallel those from a telephone survey reported
by Boruch and Cordray (1980). That survey, conducted as
a part of their larger appraisal of federal program
evaluation, indicated that in larger districts (defined
as those with enrollments of 25,000 and above), 1.6
percent of Title I allocations went to evaluation.

   Webster and Stufflebeam (1978) surveyed urban
districts nationally to gather descriptive information
about the practice of evaluation in large school
systems. Although their data are not specific as to
federal program source, the indication of the variation

TABLE C-1. Program and Evaluation Budgets of LEAs, 1978-79

Federally Funded Program	Number Having Program	Program Budget ($ thousands) Low	Median	High	Evaluation Budget ($ thousands) Low	Median	High	Percentage of Program Budgeted for Evaluation Median
Smaller LEAs (number = 30)								
ESEA Title I, for disadvantaged students	21	104	700	1,214	0	13	92	1.5
ESEA Title I, for migrant students	6	46	150	423	0	3	6	1.0
ESEA Title IV-C, innovative curricula	13	10	50	115	0	10	100	7.0
ESEA Title VII, bilingual programs	6	85	230	440	0	5	26	1.5
ESAA Emergency School Aid Act programs for desegregating LEAs	4	140	520	4,354	5	8	31	1.5
94-142, special education programs	13	50	110	600	0	3	43	2.0

260

	Larger LEAs (number = 25)							
ESEA Title I, for disadvantaged students	21	1,078	4,770	52,000	17	100	935	2.0
ESEA Title I, for migrant students	4	48	290	798	3	7	41	4.5
ESEA Title IV-C, innovative curricula	16	5	250	2,112	0	17	66	4.0
ESEA Title VII, bilingual programs	13	107	390	7,372	0	18	150	3.0
ESAA Emergency School Aid Act programs for desegregating LEAs	15	350	1,410	9,400	0	37	231	3.0
P.L. 94-142, special education programs	12	110	510	10,254	0	2	299	0.5

NOTE: Low and High designate the lowest and highest values, respectively, reported for each budget item by each LEA category. For each LEA having a particular federal program, the percentage of the program budget allocated for evaluation was computed. Entered in this table are the medians of these percentages budgeted for evaluation
SOURCE: Drezek et al., 1980.

261

TABLE C-2 · Title I Evaluation Budgets of 12 Southwestern
Districts

District	Total Title I Budget $	Title I Evaluation Budget $	Percent
A	no response	75,000	--
B	2,660,923	25,000	0.9
C	4,311,745	69,607	1.6
D	4,188,526	66,320	1.6
E	12,277,805	75,000	0.6
F	3,374,458	43,000	1.3
G	9,450,000	202,973	2.2
H	4,500,000	115,661	2.6
I	3,563,071	10,000	0.3
J	2,975,878	36,740	1.2
K	2,447,020	88,036	3.6
L	5,485,432	50,999	0.9
Mean	5,021,351	71,212[a]	1.4
Median	4,188,526	66,320[a]	1.6

[a]Includes only those districts reporting both Title I and evalu-
ation expenditures.
SOURCE: Doss (1979).

in the amount of federal funds available for evaluation
also parallels the findings from the later studies (see
Table C-3). As Table C-3 shows, federal funds constitute
a considerable portion of most school district evaluation
resources. This is somewhat at odds with the finding in
Lyon and Doscher (1979) that the funding sources for the
average evaluation office is 65 percent local, 18 percent
federal, 15 percent state, and 1 percent other. This
discrepancy may be related to urban differences and to
whether flow-through monies are treated as state or
federal resources.

The ranges of funding are as great as they are
primarily because of the way in which evaluation funding
is secured and secondarily because of differences in
evaluation requirements across federal programs and
across state agencies. One way to illustrate the
situation is to describe how funds for evaluation of
three different federal programs are typically secured
using the experience of one district as a focal point of
the description. The district is the Austin Independent
School District, Austin, Texas. Although procedures are
not exactly the same in other districts, there is
considerable similarity.

TABLE C-3   Funds Expended on Research and Evaluation Activities Within Large Urban School
Districts

District	Local/State Funds[a] ($ thousands)	Federal Funds[a] ($ thousands)	Total ($ thousands)	Cost Per Student[b] ($)
New York	300	10,000	10,300	9.49
Dallas	1,451	1,060	2,511	18.05
Philadelphia	1,222	1,378	2,500	9.62
Chicago	900	1,300	2,200	4.10
Detroit	1,203	860	2,063	8.63
Boston	941	650	1,591	18.37
Los Angeles	800	780	1,580	2.59
Baltimore	0	1,299	1,299	8.17
Atlanta	845	254	1,099	13.26
Dade County	402	290	692	2.89
Austin	356	318	674	11.50
San Antonio	271	300	571	9.44
Milwaukee	274	274	548	5.00
Cleveland	260	250	510	3.98
St. Louis	140	360	500	5.95
Portland	411	50	461	7.64
Seattle	350	75	425	6.77
Cincinnati	141	253	394	5.98
Fresno	210	180	390	7.10
Nashville-Davidson	226	164	390	5.00
Denver	336	0	336	4.50
San Jose	275	60	335	9.01
New Orleans	294	0	294	3.15
Fort Worth	155	121	277	3.89
Phoenix	141	120	261	6.50
Honolulu	194	67	261	5.14

251

TABLE C-3 (*continued*)

District	Local/State Funds[a] ($ thousands)	Federal Funds[a] ($ thousands)	Total ($ thousands)	Cost Per Student[b] ($)
Kansas City	150	83	233	4.15
Wichita	102	114	216	4.32
El Paso	96	98	194	8.14
Corpus Christi	151	33	184	4.49
Omaha	98	67	165	3.07
Dayton	148	0	148	3.59
Oklahoma City	105	19	123	2.57
Anne Arundel	114[c]	0	114	1.47
Orange County	40	27	67	0.80
TOTAL	13,002	20,904	33,906	

[a]These figures are self-reports. Where zeros (0) appear funds may be allocated for planning and evaluation to departments other than the main evaluation department.
[b]Student enrollment figures were obtained from the *Public Education Directory 1977-78*, published by Tomi Publications, Chicago, Illinois.
[c]This budget is probably somewhat higher since evaluation and research functions are performed by a number of different departments.
SOURCE: Webster and Stufflebeam (1978).

252

264

ESEA Title I

Title I evaluation is the largest federal program
activity in the Austin Independent School Ditrict (Austin
ISD) as it typically is in all SEA and LEA evaluation
units. LEA funds for evaluation are secured as a part of
an application to the SEA. The evaluation is developed
by the Austin ISD as one component of an overall Title I
program. The component sets out the scope of work to be
performed, identifies the personnel to carry it out, and
develops a budget for the activity. The amount of the
budget for the evaluation component is initially
established by the district on the basis of a district
policy statement that ties evaluation funding to program
size on a sliding-scale guideline. (This approach is not
typical since most agencies lack such a policy
statement.) What goes into the Title I application for
evaluation is generally affected by the attitude of the
LEAs toward evaluation, the way in which the application
content is controlled within the LEA, the evaluation
capability of the LEA, and in turn, by all those same
factors at the SEA level. In the Austin ISD, the
development of applications is watched rather closely by
both the school board and by the top district
management. Moreover, the staff of the department
handling federal program fund applications is favorable
toward research. In Austin at one time, and in many
districts today, the application content could be almost
entirely controlled by the application writer. When this
is true and when the writer is not favorable toward
evaluation, it can have considerable impact on the
evaluation capability.

Once developed, the application is negotiated by the
district program officer with the SEA program officer.
The entire application is generally under the supervision
of one SEA consultant; the SEA evaluation unit will
almost never be involved in the review or negotiation of
the application. Similarly, the district evaluation
staff will typically not be involved in the negotiation.
The SEA program officer is very unlikely to have seen the
district evaluation report from the previous year and may
well have little appreciation for the cost of
evaluation. Since the LEA program officer will likely
negotiate with the SEA program officer, the former's
willingness to support the evaluation budget will be
crucial at this point. When this kind of situation
exists, of course, the positive or negative nature of the

265

last evaluation report may well influence the LEA program officer's willingness to offer that support.

In summary, the Title I evaluation budget at the local level may be influenced by a number of political factors many of which will not favor rigorous evaluation and reporting. A better model would provide for involvement of the SEA evaluation staff throughout the application and approval process. Not only would evaluation activities get less one-sided consideration, but--more important--evaluation staff could introduce improvements into the program plans based on the results of completed studies.

Emergency School Aid Act (ESAA)

ESAA programs have been another source of considerable evaluation funding in the past, particularly for urban school districts. When the initial guidelines for application were issued, they were in many ways model guidelines for the development of high-quality educational proposals and programs. They set up criteria for scoring proposals that were based on a number of aspects of the program including the objectives and the evaluation. The forms were laid out so that the activities and evaluation should flow from the objectives. It has been interesting to watch what has happened to the actual awarding of grants in view of that model.

The Austin ISD annually goes through an elaborate process of proposal development that involves community hearings, working with an advisory group, and extensive staff involvement. The product of such extensive political input is usually a huge, uncontrolled set of small fragmented components, one of which is evaluation. In the Austin ISD the resulting product usually involves every school campus, some community outreach, and various disciplines from counseling to remedial reading. Even under normal resource constraints, an evaluator would stand in awe of trying to develop accountability measures for implementation and achievement of objectives. There are, however, some additional resource constraints that have at times made the task out of the question; they are discussed below.

After the proposals are put in final form by the LEA, they are reviewed by SEA representatives and submitted to the federal level. Until 1979, proposals were submitted

to the regional office; now the Washington ESAA office
staff handles the projects. The ESAA office customarily
brings reader panels in to review the proposals. These
readers try to apply the criteria set up in the ESAA
application process to the proposals. These readers are
often ESAA program officers from other LEAs and from
SEAs. Again, these readers are unlikely to have any
knowledge of evaluation. Neither readers nor program
officers often understand the sophisticated set of
criteria originally established for ESAA. For example,
the original guidelines called for awarding points on the
basis of well-developed objectives. Specific percentages
were mentioned as desirable. At least regionally this
was eventually interpreted as "the more percentages, the
better." This eventually led to such meaningless
objectives as "10% of a 10% sample of high school
students will score 75% on a measure of involvement"!
Our office was told at one point that a comparison based
on a significantly higher performance of a program group
over a control group was unacceptable.

   In the early 1970s, the Austin ISD did try meaningful
evaluation in ESAA programs several times. We had
budgets of as much as $84,000 for a program with a budget
of $840,000 for the ESAA bilingual component. (At one
time, Austin had three large ESAA programs: basic,
pilot, and bilingual, so that the annual ESAA program
budgets totalled almost $2 million.) More recently, as
the impact of Austin's last court order on desegregation
declined, funding declined as well, and evaluation
budgets fell more drastically than program budgets.
Thus, for the last three years, the evaluation/program
budgets for ESAA basic (the only component remaining in
Austin by last year) have been respectively: $3,000 and
$163,970 in 1979; $12,000 and $414,255 in 1978; and
$5,400 and $488,900 in 1977. The drastic decline in the
evaluation budget from the early years to 1977 was due to
a regional, or perhaps national, interpretation of the
legislation that a set-aside of 1 percent for national
evaluation was a limit on local evaluation as well. Of
course, there is a considerable difference between what
can be done with 1 percent nationally and what can be
done with 1 percent of a small local budget. Any true
evaluation of local ESAA became impossible even when that
evaluation was merely the mandated measurement of
objectives set out by the SEA. Such objectives had to be
carefully written around what could be measured by using
existing district data, whether they had a strong

2**7**

relationship to the program activities or not, since ESAA
funds cannot be used to purchase tests.

However, by that time we had learned that ESAA grants
were generally going to be funded late and that,
consequently, program implementation would lag badly. We
could predict that results from the program would not be
significant. In addition, for some reason, Austin has
consistently been placed on hold by the Office of Civil
Rights for the receipt of ESAA funds, and programs do not
begin until after school begins--too late for hiring good
staff or developing good programs. ESAA seems to this
writer a model for how not to do federal programming.

ESEA Title VII Bilingual Ed ation

A third type of evaluation experience came under ESEA
Title VII bilingual education. For this grant the Austin
ISD submitted a 5-year proposal directly to the Office of
Education in the spring of 1976. It had been initially
reviewed by the Texas Education Agency. Although it is
customary for Title VII to require third-party
evaluation, the Office of Education program officer
working with Austin at that time was uniquely interested
in true research and was convinced that the
organizational placement of the Austin ISD's Office of
Research and Evaluation, reporting directly to the
superintendent and the board, did indeed make its program
independent. The officer believed that it could function
within the district and with the Office of Education as a
third-party evaluator and that it could produce work of
value to bilingual evaluation in a special way. This
5-year grant has permitted a longitudinal evaluation of
the district's bilingual education effort that has
provided distinctive information and has had a real
influence on the conduct of the bilingual program in the
school system. It constitutes one of the few
longitudinal evaluations of bilingual program students in
the country; the findings have been disseminated through
a national conference held in August 1980 with the joint
support of the National Institute of Education, the Texas
Education Agency, the Austin ISD, and a number of other
agencies.

The budgets during those years have been adequate to
permit a fairly high-quality evaluation that focused in
its early years on implementation and process evaluation
and later on the longitudinal outcomes. The first-year

2⟨8

(1976) evaluation budget was $88,168 with a program
budget of $845,908; the fifth-year (1980) evaluation
budget was $60,094 with a program budget of $563,000.

Summary

Federal program evaluations are secured by LEAs through
applications to one of three agencies: SEAs, regional
offices of the Department of Education, or the Washington
office of the Department. The LEA application to the SEA
is typical of Title I, Title I migrant, and Title IV of
ESEA; of certain vocational programs; and of certain
special education programs. Generally, these grants are
"flow-through" monies: that is, funds are allocated to
states based on such factors as census information about
the number of low-income students in a state. In some
cases, the state in turn allocates set funds to districts
based upon similar census information. In other cases,
such as with Title IVC for innovative programming, funds
are allocated at the state level on a competitive award
basis. ESAA grants have come through the regional office
in the past and more recently through Washington. The
ESAA Title VII bilingual grant is typical of awards
secured directly from Washington. These are generally
competitive although there is little doubt that political
factors weigh heavily in the decisions. For example, the
size and importance of bilingual populations within a
state and city seem to be important factors in decisions
on Title VII.
    Methods and sources of funding are constantly changing
at every level, as indicated by the shift in ESAA funds
from the regional office to Washington. Other funds may
be shifted from Washington to the SEA. Each such change
results in changes in the procedures for securing funds.
Rare is the evaluation office in which staff remains
sufficiently aware of these changes and of new sources of
funds to be sure that all the available resources for
evaluation are tapped.
    At the SEA level, funding for evaluation is typically
a portion of the funds set aside for administrative
costs. This arrangement tends to pit the evaluation unit
at the SEA level against the program administration for
resources. The SEA policy on evaluation may well be the
determining factor in how much is allocated to
evaluation. Some states, particularly large ones such as
Texas, will also have regional units or service centers.

The service centers' role in federal program evaluations
is typically not large. They may perform evaluations for
small districts on a contracted basis. In some cases
they compete with LEAs for grants, such as Title IVC, and
their evaluation activities on those grants will parallel
those of the LEAs; their evaluation reports will be
provided to the central SEA just as those by the LEAs.

Regardless of the source of the funds, it should be
clear that the size and content of the evaluation
components of all programs are much influenced by program
officers at local, federal, and state levels. In the
Drezek and Higgins (1980) survey, only 21 percent of
state and local evaluation units reported that evaluation
costs were allocated on the basis of a fixed percentage
(see Table C-4). Therefore, it is important to note that
the control of the budget by program officials is likely
to have a real impact on the content and potential
credibility of evaluations.

WHO DOES EVALUATIONS?

In most states certification standards are applied to
personnel in federal programs. For example, a counselor,
administrator, or supervisor must be certified to fill
those roles in Texas. In general, evaluators are not
certified and no standards are applied to the personnel
filling the role of evaluator. In some LEAs and SEAs,
the federal program director or coordinator may bear full
responsibility for evaluation, and even in agencies with
substantial evaluation units, small federal evaluations
may be done by program staff. Typically, when program
staff are given the responsibility for evaluation, they
will have neither training nor experience in evaluation
methodology, measurement, or statistical analysis. The
author has observed many small school districts in which
the person charged with Title I program evaluation is a
reading teacher brought directly from the classroom, not
only with no training in evaluation, but also with a weak
background in mathematics.

By contrast, in some states and for some programs,
third-party or contracted evaluations are the rule. The
qualifications of the personnel in the contracting
agencies will generally vary as much as those of the
staff in the LEAs. In addition, although third-party
evaluations are supposed to ensure a lack of bias, the
contractor sometimes has an eye on future contracts and

TABLE C-4 Methods Used to Determine Program Evaluation Budget in Each Type of Agency

Method	Smaller LEAs (number = 28) percent using method	Larger LEAs (number = 24) percent using method
A roughly fixed percentage of program costs is used.	25	21
An amount is determined by the scope of evaluation work.	54	58
As much as possible, since sufficient amount is seldom received.	25	4
Other method. Examples included "all three of the above," "no fixed rule," need to consider salary levels of available staff.	21	21

NOTE: Some respondents indicated using more than one method. The number of people who indicated that they used a particular method was usually slightly larger than the number who went on to report the actual percentage, or range of percentages, used.
SOURCE: Drezek et al. (1980).

may well be gentler in approach than internal evaluators who are permanent staff.

Finally, in many districts and particularly in the large urban systems, well-trained and sophisticated evaluators with doctorates in research and evaluation carry out evaluation tasks. Within those districts having research and evaluation units with such staff, evaluator competencies are reported to be at a fairly high level in most traditional evaluation and statistical areas. In the Webster and Stufflebeam survey (1978), for example, competencies in areas such as multivariate inferential statistics, measurement theory, and experimental design were estimated by departments to be about 3.5 on a scale of 1 to 4 where 4 is "advanced competency." In newer methodologies such as bayesian

analysis and econometric applications, however, the
estimates were much lower, ranging from a low of 1.54
where 1 is "no familiarity."
Despite the rather optimistic estimate of the
competencies existing in the larger evaluation units, the
author feels that even in this area there are
considerable problems both in preservice preparation of
evaluation personnel and in-service training for current
staff. These problems deserve serious consideration.

Preservice Evaluator Training

The competencies required in evaluation are many and
varied. Boruch and Cordray (1980, Ch. 4:1) point out the
misconception that any one evaluator ever could or should
have "all the skills necessary for any evaluation
effort." It is thus obvious that any evaluator training
program has to involve choices among the many types of
skills that evaluators may eventually need. The training
that most applicants have evidenced to the author falls
short of the minimum requirements needed for a public
school evaluation office in three fundamental ways. The
applicants lack the degree of statistical and computer
programming skills needed; they do not have the
certification required by many public schools; and they
do not have adequate preparation for dealing with the
organizational and political context of the public
schools. Over the years the author has found that it is
possible to help bright candidates pick up the latter
skills and even to provide rather quickly a necessary
understanding of the evaluation task as opposed to the
research task, but the minimum statistical and computer
skills are an absolute entry necessity. Many of the
current "evaluation training" programs focus on
evaluation theory, but fail to provide adequate training
in the fundamental skills. Even though many school
systems do make it possible to hire evaluation staff
without teacher or administrator certification, few will
permit the evaluator without those credentials to move to
administrative positions in the evaluation office. Many
evaluators do not even realize that such credentials are
needed although in many cases it might have been
relatively easy for them to pick up such certification as
a part of their graduate programs.
There are a number of steps that might lead to better
preservice training that could be taken by the Department
of Education or Congress. For example:

* Designers of preservice training programs
receiving federal support might be required to involve
in-service evaluators;
* Federal support might-be given to graduate
training programs that contain provisions for field
experience and internships in an LEA or SEA; ˙
* Field experiences in an LEA or SEA could be
. offered early in a training sequence, thus providing
exposure to requirements in those settings;
* Support might be given to interchanges between
university and SEA or LEA evaluation staff of one or two
semester lengths. so that university programs do not
become too insular.

In-Service Evaluator Training

Since a preservice program cannot possibly give an
evaluator all the skills that will eventually be needed
and since many practicing evaluators do not presently
have even the minimum skills, better in-service training
opportunities for evaluators are desperately needed.

Many conditions limit practicing evaluators from
maintaining and increasing their skills at the present
time. Public school evaluation is an all-consuming
role. An evaluator works 12 months, with summer bringing
the heaviest work load; because resources are often
inadequate, the workday and workweek are far longer than
those of the average worker. Therefore, once an
evaluator is on the job, there simply is not sufficient
time available to renew or enhance skills. Turnover of
evaluation staff is high: the Austin ISD loses 25
percent of its evaluation staff (15 senior and 20 junior
professionals) every year. Perhaps there is such high
attrition not only because of the time demands but also
because-evaluation is an emotionally difficult field.
The constant negotiations necessary have been described
in several chapters of this report, but inevitably, many
practicing educators fear and dislike evaluation and
resent the power that comes with evaluation information.
The evaluator must deal with those negative feelings on a
daily basis. At the same time, the professional rewards
for an evaluator in an LEA or SEA are few. The social
science research community tends not to esteem evaluation
work very highly, and evaluation specialists in
-universities give limited recognition to work carried on
elsewhere. Thus, there is little in an evaluator's

environment that even encourages staying in the field,
not to mention participating in additional training if it
were available. In fact, however, additional in-service
training is really not even available. There are such
things as AERA presessions, and the Austin ISD staff
regularly participate in those. There are a few
week-long university sessions offered during the summer,
but summer is the busiest time of the year for an
evaluator. (The only time with any slack at all in the
Austin schedule is November, December, and January.) And
when the evaluator does participate in any of these
activities, they tend to be piecemeal and disjointed.

In the face of such a grim diagnosis, are there things
that could be done to improve in-service learning
opportunities for evaluators? Yes, but most of those
things will be very expensive, such as:

- Post-doctoral residential programs in which
evaluators return to university training for a semester
or two;
- The exchange programs between university and
LEA-SEA staff mentioned above would be beneficial to the
evaluator as well as the university programs;
- Special project assignments at the federal level
with built-in training by resident staff;
- Special training sessions planned and offered on
a sequential basis at times favorable to LEA and SEA
evaluation schedules;
- Visiting scholar programs such as those already
being offered on a limited basis by the Center for the
Study of Evaluation.

In addition to such formal efforts, however, much can
be done on an informal basis to encourage an evaluator's
professionalism and to provide incentives for learning.
The author has received enormous benefits in that sense
from the network membership established through Division
H of AERA and the Directors of Research and Evaluation.
The evaluation report awards given annually by Division H
were created to provide recognition for evaluation work.
The new Journal of Educational Evaluation and Policy
Analysis may provide a publication forum for evaluators.
Recently, the Title I technical assistance center for the
region serving Texas has brought together the Title I
evaluators from large cities to form a network
relationship for this region. Such networks could be of
considerable help in increasing the professionalism of

federal program evaluation staff related to such other
programs as Title VII, special education, and career
education.

WHAT HAPPENS IN EVALUATIONS?

Compliance activities probably predominate in the
majority of federal program evaluations at both the SEA
and LEA levels. In many SEAs this may be almost the sole
preoccupation. They will design annual report documents
to gather information from LEAs, gather such information,
and provide it in turn to federal offices. They are
likely also to conduct or participate in monitoring
visits to LEAs to check fiscal and program plan
compliance. Only a few states currently attempt more
substantive studies designed to influence state plans for
the use of federal program funds or to evaluate the
effectiveness of program activities, although the
activities in several states are noteworthy.
   At the local level, the first priority activities for
the evaluation unit also may well be data collection
relative to compliance. For example, one of the largest
aspects of Title I evaluation may be the collection of
data on low-income enrollments by campus, the
identification of students eligible for service based on
low achievement, and locating students in nonpublic
schools or who have dropped out. Until the advent of the
Title I models, much of the reporting involved little if
any analysis. Similar activities and numbers are
fundamental in most federal program evaluation efforts.
   After these compliance or record-keeping types of
activities, the measurement of performance relative to
set objectives is probably the next most typical
evaluation activity. Great variety exists across
programs in the type of objectives established. I have
already touched on those used in ESAA programs; other
types may range from achievement outcome objectives to
service objectives based on the number of participants
served. The survey of Title I programs in the Southwest
mentioned earlier (Doss 1979) yielded information that
demonstrates both the nature of Title I objectives in
reading and a feel for the variety of test instruments
used. (Some representative samples are shown in Table
C-5.) Boruch and Cordray (1980, Ch. 5:11-12) have
appropriately criticized such objectives as arbitrary and
insufficient as standards for evaluation. After far too

TABLE C-5 Reading Achievement Objectives in Southwestern Title I Programs

District	Grade(s)	Testing Pattern	On Level?	Test	Expected Gain
A	2-8	Spring to Spring	Yes	MAT	DISTAR Reading I Program--65 percent will show a gain of 0.6 mo./mo. of instruction. DISTAR Reading II Program--60 percent will score at the 2.9 reading level. DISTAR Reading III Program, High Intensity Program, and Reading Skills Program--60 percent will show a gain of 1 mo./mo. of instruction.
C	1	Spring	Yes	CAT	55 percent will show a gain of 0.1-0.6 mo./mo. of instruction. 30 percent will show a gain of 0.7-1.0 mo./mo. of instruction. 15 percent will show a gain of 1.1 mo./mo. of instruction or more.
	6-7	Spring	No	Local Criterion-Referenced Test	60 percent will attain 50 percent of grade level reading objectives. 30 percent will attain 51-60 percent of grade level reading objectives. 10 percent will attain 61 percent or more of grade level reading objectives.

2.0

F	1	Fall to Spring	Yes	CAT	For 80 percent of the Title I partici-pants it is expected that the mean posttest stanine will be greater than the mean pretest stanine. For the re-maining 20 percent, the posttest stanine will remain the same as the pretest.
H	1-6	Fall to Spring	Yes	MAT	75 percent will gain at least three per-centile points.
	7-8	Fall to Spring	Yes	MAT	70 percent will gain at least three per-centile points.
I	K-8 9-12	Fall to Spring	Yes	MAT	An NCE gain that exceeds zero.
		Sept., Nov., Jan., Mar., May	Yes	G-M	An NCE gain that exceeds zero.
K	1	Spring Only	Yes	CST or BRST	Will make progress in reading readiness.
	2-8	Fall to Spring	No	CAT	Will show an NCE gain from pretest to posttest on a composite reading score.

SOURCE: Abridged from Doss (1979).

265

277

much experiece in dealing with objectives in evaluation,
I have concluded that they may be a great tool for
planning, but they are a poor tool for evaluation.

Only in a few instances are substantive, long-range,
or cumulative effects of federal programs examined. As
we in Austin ISD have struggled with federal program
evaluation over the years, we have become convinced that
such evaluation produces the best information and leads
to the best utilization.

An interesting trend in the last few years has been
toward what have been called "interpretive analyses,"
such as: Impact of Title I: A Decade of Progress (Moore
and Turner 1976); Limitations of a Standard Perspective
on Program Evaluation: The Example of Ten Years of
Teacher Corps Evaluation (Fox 1977); Evaluation in the
Seventies: What We Have Learned About Program
Development and Evaluation (Holley 1977). These reports
try to bring together information gained from discrete
evaluation efforts either across years or across programs.

HOW ARE EVALUATIONS REPORTED?

Evaluations are reported in a number of ways, both formal
and informal. There is probably less uniformity from
district to district in reporting than in either
budgeting or in activities. Again, it may be
illustrative to use the Austin ISD procedures as the
center of this discussion of reporting. ESEA Title I
involves the most elaborate reporting and is therefore
used as the example. The flow of information is charted
in Figure C-1.

The school year in Austin runs from July 1 each year
to June 30 the following year. Austin's major reports
come at the end of the year and the month of June is a
hectic, full month of analysis, interpretation, and
report writing. As for all Austin ISD evaluation
projects, the Title I evaluation staff prepare a final
technical report and a 15-page final report. The
technical report consists of appendices covering each
data collection effort. It is long and voluminous; only
a few copies are produced. The 15-page report goes into
a book called Findings Volume. The short report is the
major communication vehicle about the project. It covers
the essential results first, then describes the project
and the evaluation and provides some discussion of the
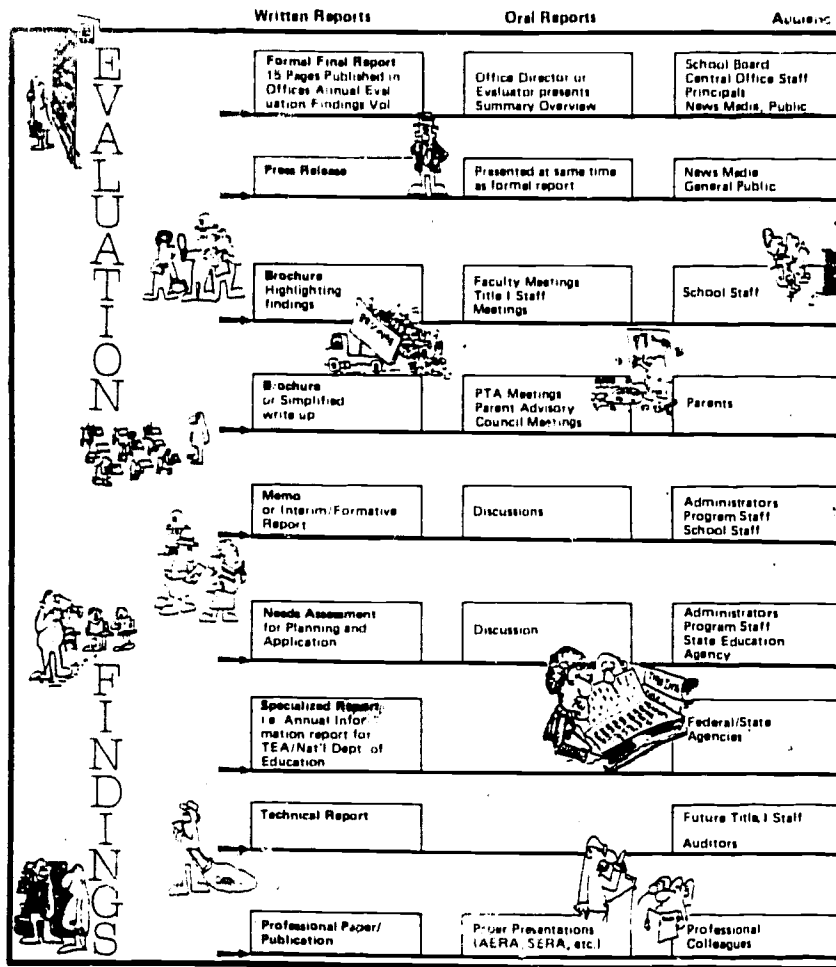results. This short report evolved from our growing

FIGURE C-1 Evaluation reporting for ESEA Title I in Austin, Texas.

experience that reports longer than 15 pages never got read at all. In addition, Title I staff must complete an AIR report--an annual information report--to the SEA. This is a form containing numbers, analysis of the achievements of various components, and a space to indicate changes to be made as a result of the evaluation data. The Texas Education Agency has put considerable effort into improving this reporting form over the years in an attempt to encourage good evaluation and utilization.

The AIR report is signed by the superintendent and submitted to the Texas Education Agency. It is not reviewed by the school board primarily because the board will receive the Findings Volume, which contains the same results but in the usual district format. The format is of concern because, given the limited time available for the presentation and discussion of evaluation results, it is important not to have to expend time or effort to explain differing formats. Soon after June 30, which is the annual deadline for the completion of final reporting, a session with the school board to review all results is held. Thereafter, all reports become public information and freely available. Copies of both the technical and final reports are placed in the board office, the district's professional library, and the Office of Research and Evaluation. Presentations of the results are then arranged early in the school year for principals, instructional staff, and various other groups. All of these formal presentations, however, are not nearly so important as the informal discussions that subsequently occur. Knowledge of important findings relevant to a specific instructional supervisor or administrator may be shared over coffee or lunch. In particular, findings may be reviewed during planning sessions for particular programs or activities.

A follow-up reporting activity for the past few years has been a short brochure summarizi Title I results for teachers and parents. Results are also mentioned in newsletters.

Another critical reporting perio Title I comes during the early part of the calendar year. It is the needs assessment for the preparation of the next year's program plan. This assessment reports data about where students will be and what achievement levels are. From this report, Title I schools for the following year will be designated and cut-offs for eligibility will be established. The report is mainly for in-district use,

but an abstract is provided to the board and the volume
itself placed in the board office. Then it becomes
public information and is available to the community. It
is often used by other agencies in the city in their
preparation of proposals for funds.

Thus, all reports prepared about Title I are available
for public scrutiny. I do not know whether this is
common practice around the country. Although certainly
in Texas all submitted reports are public documents and
thus available to all, many districts do not make the
availability of reports well known. Also, reports are
not always submitted to school boards. This may either
be because the superintendent wishes to keep the reports
internal or because the board is not interested in them.

WHAT IS THE IMPACT AND USE OF EVALUATION?

Given the picture described above, it would hardly be
surprising if the impact and use of evaluation at the
state, regional, or local level were difficult to trace
or document even if we had good procedures for doing so.
Much of the current literature on utilization seems to
conclude that utilization does occur, but that it takes
diverse and difficult-to-trace routes. This writer's
subjective observations concur with that conclusion. As
a program officer from another Texas district told a
group recently, prior to the advent of federal programs
you could walk into a school and ask how well the
students were performing and never get anything but
subjective answers. Now schools all over the state know
precise levels at which students, schools, and districts
are performing. Sometimes they can even tell you why the
levels are what they are. Because federal programs are
now so pervasive, we often fail to recognize just how
great their impact on the conduct of schooling has been.
It has been clearly demonstrated in Texas that where
evaluation produces useful results, they do get used in
program design. Eventually.

This is not to say that impact and utilization are
what one would wish. It is of major concern to this
writer that the effects of evaluation are only a fraction
of what they might have been if the resources that have
been available had been more carefully guided and
targeted. However, evaluation has been an innovation and
we are only now learning m᠁ny of the things we needed to
know about its implementation.

The fundamental lack of evaluation information that
could contribute to the overall design of better programs
is one of the most serious handicaps to extensive use.
It has been a particular idea of this writer that on
programs such as Title I or Title VII, for which we are
expending rather large sums in local evaluations, we
might find better ways to capitalize on that evaluation
effort. If evaluations of compensatory programs were
coordinated in even a minimal way, how much richer our
evaluations might have been. For example, teachers's
aides and other instructional aides are commonly used in
various compensatory programs, yet, their effectiveness
has been examined only in an incidental way in a few
evaluations. What many of us have found in those
examinations has, however, been disturbing. The data are
not complete enough for conclusive statements about the
effectiveness of aides; it might have been if a larger
number of school districts had examined how aides were
being used and what the effects were. The use of time is
another important factor that affects outcomes that some
of us have stumbled on in our evaluations. Again, data
across a large number of districts collected through
careful observation studies would be far better than
estimated numbers on every child in Title I filled in
capriciously from district to district. What are some of
the ways such an idea might be accomplished? A number of
ways can be imagined, varying from fairly indirect to
direct and controlled.

In Texas, for example, a number of urban districts
have regular meetings of their superintendents,
curriculum staff, and evaluators. These meetings have
led to the sharing of information among each group. The
meetings of the evaluation group, the Joint Urban
Evaluation Council, has resulted in similar studies on
several topics in the different cities. Measures and
reports have been exchanged. Support for the national
directors of research and evaluation (DRE) group, which
now meets annually for one day prior to the AERA meeting,
to have more frequent meetings might have similar results
at the national level. Such a forum could be used for
the Department of Education to present a set of critical
issues in compensatory education and possible alternative
evaluation designs to address these.

The Title I technical assistance centers (TACs) might
also be given the task of the informal encouragement of
such efforts as they work with school districts. In
informal discussions with one TAC center evaluator, I

discovered that such encouragement might already be
happening. Another role that the TACs could play that
would contribute in the same sense as the regular DRE
meetings would be that of bringing the Title I evaluators
together on a regional basis. Although mentioned already
as a route to improved in-service training for
evaluators, it could also be a stimulus to shared designs.

The fundamental lack of important evaluation
information that could contribute to improved programs
and failure to coordinate information that does exist are
not the only handicaps to utilization, however. There
are other factors. First, federal programs in general
tend not to be of high concern to most local school
boards and administrators. This can be interpreted more
as a matter of time available and priority than as a lack
of interest (Holley 1980). The federal funds in the
Austin ISD, for example, are currently about $5 million,
but this is only a fraction of the total district
operating budget of well over $100 million. While this
ratio is smaller than for many districts, it is still
fairly representative. Austin has had far better
attention to federal programs and their evaluation since
the Board of Trustees adopted as one of its top
priorities to improve the achievement of low
socioeconomic and minority students. The board adopted
this priority based on evidence of the enormous deficit
in the achievement of those students relative to the
total student body and because they represent a growing
proportion of the student body. With this general
priority for these students in the district, federal and
state compensatory programs come into focus as one of the
major resources for achieving district priorities. The
Department of Education may find that strong federal
program evaluation coincides with strong district
evaluation.

Another obstacle to the use of federal program
evaluation information is the lack of recognition of
dissemination needs. Typically, an evaluation is
coterminus with a program grant. For example, when the
Austin ISD recently applied for a 2-month extension of
its 5-year study of the Title VII bilingual program in
order to provide for more extensive dissemination, the
request was denied despite the fact that no new monies
were requested. Had our office not felt the evaluation
results were so important that we devoted nonfederal
resources to dissemination efforts and continue to do so,
much of the value of an important evaluation study would
have been lost. Such constraints mean in many cases that
no dissemination of findings ever occurs.

283

Still another barrier to dissemination lies in the
area of communication. Anyone who has worked
consistently in evaluation realizes that the time
available for communication of evaluation results is
never adequate. In a large district with many competing
communication needs and with many evaluations, this is a
severe problem. Efficient evaluation units develop
communication strategies that permit the telescoping of
information through shorthand forms for reporting. Since
the data that will have impact at one level of the system
are not the same as those that will have impact at
another, the information has to be transmuted innumerable
times before dissemination is accomplished. Resource
needs for this effort may well not be recognized. Thus,
the improvement of utilization must come both through
better evaluations that produce more useful information
and through better dissemination and promotion techniques
on the part of the evaluation staff. Both efforts need
better recognition and better support from Washington.

CONCLUSION

Variation is the theme around which this paper is
written, and surely that theme has been demonstrated.
Complexity of relationships may have emerged as a major
subtheme, however. Figure C-2 lays out some of the
funding, reporting, and advisory relationships as they
appear from the experience of the author. Each year the
complexity seems to increase with a concurrent decrease
in the flexibility available to the LEA.
    Every increase in complexity has tended to bring
additional reporting demands to the LEA. Ultimately, the
bulk of that reporting burden falls on students,
teachers, and principals. To the extent that such
reporting has moved beyond their central concerns, it
becomes meaningless bureaucracy. This in turn has two
serious side effects. There will be an increased dislike
and disrespect for "evaluation," and there will be a
decreased willingness to hear and utilize evaluation
results.
    Both Congress and the Department of Education would be
wise to consider such effects in designing national
evaluation requirements and systems. Ultimately, the
most successful evaluation of federal programs will be
that which leads to programs that are winners--winners
for both students and staff.
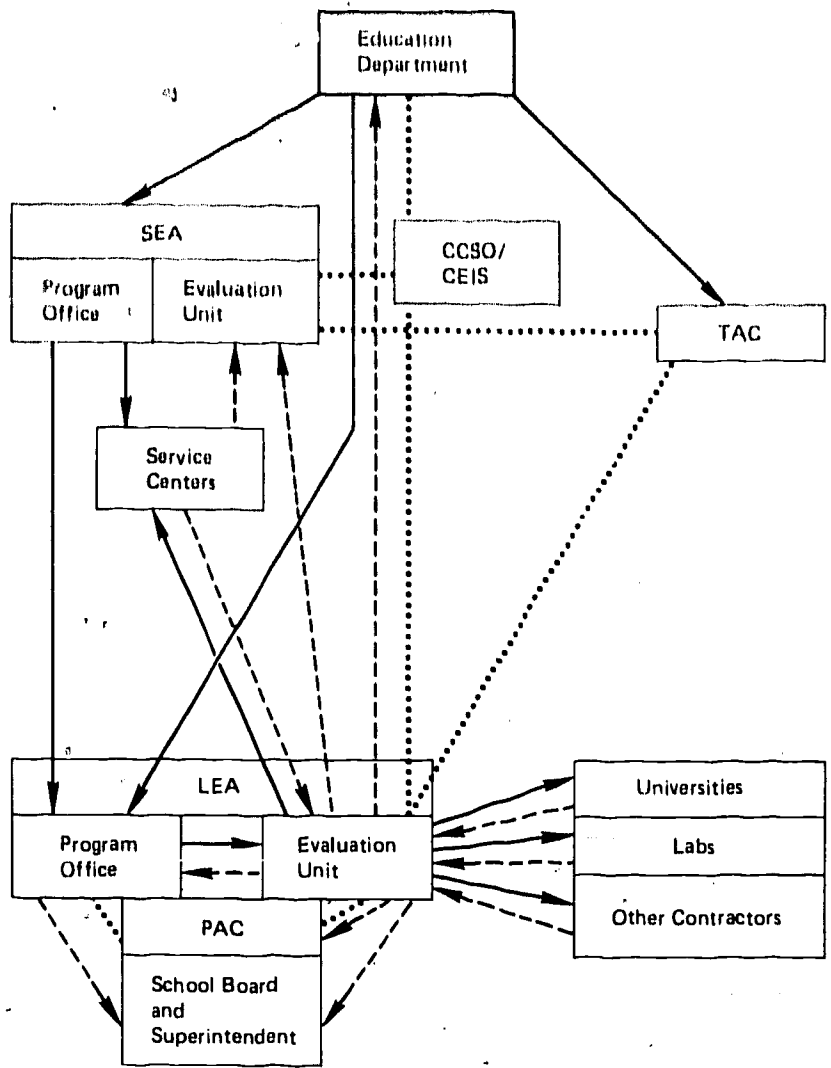
Funding Relationships ───────►
Reporting Relationships ── ── ──►
Advisory Relationships ••••••••••

SEA = State Education Agency
LEA = Local Education Agency
PAC = Parent Advisory Committee
TAC = Technical Assistance Center
CCSO/CEIS = Chief State School Offices/
  Committee on Evaluation and
  Information Systems

FIGURE C-2   Relationships in LEA evaluation of federal programs.

REFERENCES

Boruch, R.F., and Cordray, D.S. (1980) An Appraisal of
Educational Program Evaluations: Federal, State, and
Local Agencies. Prepared for the U.S. Department of
Education, Northwestern University.

Doss, D. (1979) Correspondence With Other Districts.
Appendix BB in ESEA Title I Regular Program 1978-79,
Final Technical Report. Publication #78.61. Austin,
Texas: Office of Research and Evaluation, Austin
Independent School District.

Drezek, M., and Higgins, P.H. (1980) Current vs. Ideal
Procedures for Determining Educational
Program-Evaluation Budgets: A Survey of School
Evaluators. Paper presented at Division H Task Force
meeting of the American Educational Research
Association annual meeting, Boston.

Fox, G.T., Jr. (1977) Limitations of a Standard
Perspective on Program Evaluation: The Example of Ten
Years of Teacher Corps Evaluations. Madison, Wis.:
University of Wisconsin.

Holley, F.M. (1977) Evaluation in the Seventies: What
We Have Learned about Program Development and
Evaluation. Publication #75.59, ED 141-940. Austin,
Texas: Office of Research and Evaluation, Austin
Independent School District.

Holley, F.M. (1980) Evaluation Utilization: Is It
Easier to Move a Mountain than a Molehill? Paper
presented at the American Educational Research
Association annual meeting, Boston.

Lyon, C.D., and Doscher, L. (1979) Executive Summary:
Evaluation and School Districts Report. Los Angeles,
Calif.: Center for the Study of Evaluation,
University of California.

Moore, T., and Turner, W.E. (1976) Impact of Title I: A
Decade of Progress. Wichita, Kan.: Department of
Program Evaluation, Wichita Public Schools.

Webster, W.J., and Stufflebeam, D.L. (1978) The State of
Theory and Practice in Educational Evaluation in Large
Urban School Districts. Paper presented at the
American Educational Research Association annual
meeting, Toronto.

D

Individuals Interviewed and External Participants in Committee Meetings

JOEL ANTHONY, Office of Administration, Management, and Budget, National Institute of Education

KEITH BAKER, Office for Planning and Management, U.S. Department of Education

L. VAUGHN BLANKENSHIP, Director, Division of Applied Research, National Science Foundation

LOIS-ELLIN DATTA, Associate Director for Teaching and Learning, National Institute of Education

JANE L. DAVID, President, Bay Area Research Group, Palo Alto, California

PRISCILLA (PAT) E. DEVER, Administrative Officer, Program Evaluation, U.S. Department of Education

JOHN W. EVANS, former Assistant Commissioner for the Office of Evaluation and Dissemination, Office of Education, U.S. Department of Health, Education, and Welfare

JOHN GABUSI, Assistant Secretary for Management, U.S. Department of Education

EDWARD B. GLASSMAN, Office for Evaluation and Management, U.S. Department of Education

WILLIAM A. HIGHTOWER, Human Resources Division, U.S. General Accounting Office

HOWARD F. HJELM, Director, Division of Research and Demonstration, Office of Vocational and Adult Education, U.S. Department of Education

BOBBY R. HOOVER, Human Resources Division, U.S. General Accounting Office

SAMUEL W. HUNT, Staff, Appropriations Committee, U.S. Senate

*Affiliations of individuals at time of interviews

ALBERT B. JOJOKIAN, Group Director, Human Resources
Division, U.S. General Accounting Office
JOHN JONAS, Legislative Assistant to former
Representative Elizabeth Holtzman (author of provision
in P.L. 95-561 to assess program evaluation in
education)
ROLF F. LEHMING, Dissemination and Improvement of
Practice, National Institute of Education
RICHARD T. LOUTTIT, Director, Division of Behavioral and
Neural Sciences, National Science Foundation
ROBERT J. MARONEY, Office of Evaluation and Management,
Program Evaluation, U.S. Department of Education
JOHN M. MAYS, Office of Director, National Institute of
Education
LINDA MORRA, Office for Special Education and
Rehabilitative Services, U.S. Department of Education
ELIZABETH R. REISNER, National Testing Service Research
Corporation, former staff, Office of Education, U.S.
Department of Health, Education, and Welfare
ALFRED R. SCHNUPP, Human Resources Division, U.S. General
Accounting Office
DOROTHY A. SHULER, Office of Evaluation and Management,
Program Evaluation, U 3. Department of Education
JOHN SEAL, Deputy Assistant Secretary for Evaluation and
Management, Office of Management, U.S. Department of
Education
MARSHALL (MIKE) S. SMITH, former Assistant Commissioner
for Policy Studies, Office of Education, U.S.
Department of Health, Education, and Welfare
CARL E. WISLER, Office of Evaluation and Management,
Program Evaluation, U.S. Department of Education
JOSEPH S. WHOLEY, former Deputy Assistant Secretary for
Evaluation, U.S. Department of Health, Education, and
Welfare
ROSEMARY C. WILSON, Director, Division of Follow-Through,
Office of Elementary and Secondary Education, U.S.
Department of Education
THOMAS R. WOLANIN, Staff Director, Subcommittee on
Post-Secondary Education, Committee on Education and
Labor, U.S. House of Representatives

2S9

"END OF DOCUMENT"