

FEDERAL REQUIREMENTS AFFECTING
LOCAL EVALUATION USE

James Burry

CSE Report No. 222

1984

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California Los Angeles

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Table of Contents

Factors Affecting Local Evaluation Use	2
Figure 1: Factors Affecting Evaluation Use	3
Federal Program Evaluation Requirements	4
Chapter 1 Evaluation Requirements	6
Chapter 2 Evaluation Requirements	6
Bilingual Program Evaluation Requirements	7
Education for the Handicapped Evaluation Requirements	8
Vocational Education Evaluation Requirements	9
Broad Problem Areas	10
An Ideal Scenario	12
How the Scenario Factors May Promote Local Evaluation Use	16
The Effects of Current Federal Requirements	19
Chapter 1 Evaluations	19
Chapter 2 Evaluations	21
Bilingual Program Evaluations	22
Handicapped and Vocational Education Evaluations	24
Some Observations from the Field	25
Summary	28
References	31

FEDERAL REQUIREMENTS AFFECTING LOCAL EVALUATION USE

. . . As the shift continues from the federal to the state levels in the management of education programs, the states become more, not less, accountable for them. SEAs and LEAs have become accustomed to the federal government not only requiring the evaluation of programs but also dictating methods of evaluation (Council of Chief State School Officers, 1982).

For the past four years, the Center for the Study of Evaluation (CSE) has focused on identifying the factors which affect local uses of evaluation. As defined here, the term use refers to the planned actions that result from applying evaluation information and processes to the resolution of specified problems, questions, or concerns. While evaluation can have unplanned impacts and consequences, our research has emphasized only those which are intended to take place. The analysis in this paper is further limited to use that occurs at the local program level.

Use factors comprise those characteristics or elements present in evaluation settings that potentially affect use. Our research has identified a number of such factors, and this paper focuses on one of them. Termed "written requirements for the evaluation," this factor includes the evaluation requirements accompanying federally funded programs. Two questions concern us here: how does this factor interact with other factors in the evaluation? and is this interaction likely to have a positive or negative effect on use?

After discussing what our research has uncovered about the factors affecting local evaluation use, we will describe the federal program evaluation requirements currently in effect. Drawing on our research findings,

we will then present an ideal scenario in which the factors are organized to have a high potential for stimulating local evaluation use. Finally, we will discuss the impact of federal requirements on factors promoting local evaluation use.

Factors Affecting Local Evaluation Use

Our research on evaluation use falls into four phases. The first phase, case studies of school-level program implementation and evaluation, identified some of the factors affecting use and led to an agenda for further research (Alkin, Daillak, & White, 1979). The second phase, field studies of evaluators at work, elaborated some of these factors, especially those reflecting the organizational context of the program being evaluated (Daillak, 1980). The third phase involved a user survey which elicited a picture of the kinds of decisions educational administrators make and the kinds of information they use as these decisions are shaped (Stecher, Alkin, & Flesher, 1981). In the fourth phase, we synthesized what we know about evaluation use, about the range of factors affecting use, and about the ways in which these factors can be influenced to promote use (Alkin et al., 1983). This paper draws on that synthesis.

Figure 1 following classifies and lists the factors that potentially affect the local uses of evaluation. The individual factors are grouped into three broad categories. Human factors reflect the evaluator and user characteristics affecting use: their attitudes toward and interest in the project and its evaluation, their backgrounds and organizational positions, and their professional styles. Context factors are inherent in the setting of the particular program being evaluated; they include the characteristics of that program, the requirements and fiscal constraints imposed on the

Figure 1: Factors Affecting Use

I. Human Factors

A. Evaluator Characteristics

1. commitment to use
2. willingness to involve users
3. choice of role
4. rapport with users
5. political sensitivity
6. credibility
7. background and identity
 - a. gender
 - b. title

B. User Characteristics

1. identity
 - a. range of potential users
 - b. organizational positions
 - c. professional experience levels
2. interest in the evaluation
 - a. views about the project being evaluated
 - b. expectations for the evaluation
 - c. predisposition toward the evaluation
 - d. perceived need
 - e. perceived risks
3. commitment to use
4. professional style
 - a. administrative and organizational skills
 - b. initiative
 - c. openness to new ideas or change
5. information processing
 - a. preferences for particular forms
 - b. how information is processed

II. Context Factors

A. Pre-existing Evaluation Bounds

1. written requirements
2. other contractual obligations
3. fiscal constraints

B. Organizational Features

1. intraorganizational
 - a. role of central/district office
 - b. interrelationship between unit and central/district administration
 - c. institutional arrangements
 - d. unit level autonomy
 - e. sources of information beyond evaluation likely to be in use
 - f. perceived institutional risk
2. external features
 - a. community climate
 - b. community influence
 - c. role of other agencies

C. Project Characteristics

1. age/maturity
2. innovativeness
3. overlap with other projects

III. Evaluation Factors

A. Evaluation Procedures

1. methods used
 - a. appropriateness
 - b. rigor
2. dealing with mandated tasks
3. used of a general model

B. Information Dialogue

1. amount and quality of interaction between evaluator and users

C. Substance of Evaluation Information

1. information relevance
2. information specificity

D. Evaluation Reporting

1. frequency of information provided
2. timing of information
3. format presentations
 - a. oral presentations
 - b. written reports
 - c. statistical and narrative data

The kinds of use-promoting activities mentioned above, involving project director, evaluator, and program staff, require certain abilities in and attitudes toward evaluation. It is critical that evaluation requirements do not work against either. In the case discussed here, the program's surrounding context is such that project evaluator and director, both of whom are committed to evaluation use, do not need to fear that external requirements will work against local use. Thus, they maintain their commitment at the same time as they involve other staff in the evaluation, thereby increasing the likelihood that the evaluation will actually be put to use.

The Effects of Current Federal Requirements

Now we will take a closer look at how current federal evaluation requirements compare with those described in the ideal scenario.

Chapter 1 Evaluations

As we pointed out earlier, the federal requirements for Chapter 1 program evaluations involve assessing the achievement of specified goals on the basis of locally chosen objective measures and determining whether improved student performance is sustained over a period of more than one year. These requirements, intentionally or not, may create obstacles to local evaluation use.

For instance, the "objective measures" mentioned in the Chapter 1 requirements may not be uniformly understood. Owing to historical precedent, one local program may believe that a test must be norm-referenced in order to be considered objective and may continue to use this kind of test even when it serves no relevant local purpose. If this is the case, the evaluation findings generated by use of this test are likely to be seen by program staff as having little practical value for them.

In districts meeting the kinds of problems outlined above, that is, where testing issues are difficult to resolve, where there is limited expertise to consider technical adequacy and local relevance, there is likely to be some negative effect on factors promoting use: local anxieties and sense of risk may dominate; program staff may believe their questions and concerns are receiving insufficient attention; the evaluator's credibility is likely to suffer; local commitment to evaluation use and perception of local usefulness will decrease.

Chapter 2 Evaluations

The current requirements for Chapter 2 evaluations permit much greater flexibility and decentralization than were allowed in the past: they simply instruct state and local agencies to keep whatever records and information that might reasonably be required for fiscal audit and program evaluation. Of course, the juxtaposition of "fiscal audit" (often seen as potentially punitive) and "program evaluation" may itself have a negative effect on local use. But Chapter 2 regulations can affect local evaluation use in other ways.

Some districts may opt to conduct no program evaluation. Others may keep records of whatever tests or other measures were administered but make no effort to analyze and use their results. Still others, accepting the spirit of the requirement, may try to evaluate their Chapter 2 funded activity or activities. In doing so, they face several possible problems.

For example, a district operating a Chapter 2 program that emphasizes only one of the authorized content areas may encounter any of the barriers to uses already mentioned. On the other hand, a district that commits Chapter 2 funds to several of the authorized areas not only faces the

normal problems but may also encounter another dilemma if it tries to evaluate all of its Chapter 2 activities: namely, how to conduct evaluations of several content areas while maintaining a balance among them, and build use potential into each.

For instance, the staff of one activity may feel that it is competing with other staffs for resources, and may fear that it will not get a fair hearing. Further, if one person is selected to evaluate all the activities in operation, his or her credibility may be suspect; but if each authorized activity has its own evaluator, administrative and management conflicts are likely to arise. These problems, to the extent that they limit efforts to stimulate use-conducive factors, can have a negative effect on the evaluation and its uses.

Other problems may arise because of the varied nature of the activities that can operate under Chapter 2. Some content areas, such as school library services, are candidates for resource expenditure monitoring. Other areas, such as consumer education and ethnic heritage, are more amenable to evaluation. However, consumer education, which primarily emphasizes cognitive achievement, requires a different kind of evaluation than an ethnic heritage program which is likely to emphasize student attitudes.

In short, it seems doubtful that a district will be able to mount simultaneous evaluations of the kind suggested in the scenario without more careful guidance than is presently available under Chapter 2.

Bilingual Program Evaluation

Because the language of current federal regulations governing bilingual education seems to stress the technical aspects of evaluation,

districts often assume they will be penalized if they do not follow classic canons of methodology.

In this context, the regulations use technical terms -- such as "design," "data collection and analysis procedures," "comparison procedures ... to estimate performance," "statistical comparison," "sampling procedures" -- which are intended as suggestions about how a local program might be evaluated. But the people involved in local bilingual programs see them not as suggestions but as federal requirements (or, at the very least, federal preferences) and so spend great amounts of time and other resources trying to accomplish what they feel is expected of them (Burry, 1979).

In our ideal scenario, which centered on a bilingual program, there were few technical "suggestions" about how the evaluation might proceed. Rather, the overall tone of the requirements suggests agency concern with ensuring the likelihood that the evaluation, beyond answering agency questions, will also address the local program's evaluation needs and range of information uses. And these uses are typical of the interests of bilingual program staff -- improving instructional materials, modifying classroom practices, developing a harmonious relationship with the community. Further, the presence of full-time evaluation expertise will help ensure that these program interests receive a balanced treatment, balanced in terms of how the evaluator distributes his time and talent among them, and balanced in terms of the attention they receive in relation to external interests and information needs.

In the ideal scenario, then, we encounter one evaluation designed to meet a variety of needs, one of which is a direct focus on local uses. However, to get relevant information in the appropriate form and at the

right time, many bilingual programs mount parallel evaluations. They conduct a formal evaluation for federal purposes which involves administering (typically) a norm-referenced test, comparing the test results with some (statistical, historical) standard, estimating their program's contribution to student outcomes, and producing a final report. Because program staff, especially teachers, find this formal evaluation of little use for their purposes, their resentment toward it builds as they do their own informal evaluations to get the information they need to run their classrooms.

Handicapped and Vocational Education Evaluations

The federal evaluation requirements associated with these programs are discussed together for two principal reasons. First, handicapped and vocational education programs have traditionally not been subject to evaluation and its uses as defined in this paper. The nature of the populations served, the types of assessment devices required, the kinds of instruction provided, and the outcomes expected are such that attention is directed to the individual student rather than to the program as a whole.

Second, while they differ in the language used and specificity of direction, each program's regulations seem to culminate on the same note: reporting on the numbers of students served and on how well they are served. This note is appropriate: the desired outcome of vocational education is the successful employment of the individual student; the desired outcome of education for the handicapped is the provision of instruction and other services appropriate for the individual child. In both cases, then, the individual -- not the program -- is emphasized.

It does not follow from this, however, that evaluation information cannot be used at the program level. Indeed the regulatory language for both handicapped and vocational education contains the word "program" in association with "evaluation." But as evaluations are currently conceived and conducted, the state, which is the responsible body, seems to emphasize a regulatory function stressing compliance. In the case of education for the handicapped programs, states conduct evaluations which include site visits and which report on the number of students served in each of several disability groups. In the case of vocational education programs, states may require or suggest the use of follow-up procedures such as surveys asking graduates about their current jobs, earnings, satisfaction, and so forth; some states also send site-visit teams to observe classrooms and inspect equipment.

In neither case is there any explicit concern for the kinds of program-level evaluation aggregation illustrated in the scenario. Further, states and local districts seem to have little conception of this kind of evaluation and its uses. Should evaluation as treated in this paper be the intention behind federal requirements, then the requirements will probably have to be reconceptualized.

Some Observations from the Field

To test our impressions of the effects of current federal regulations on local evaluation use -- particularly with respect to the obstacles they create -- we conducted some interviews with the people most responsible for complying with federal regulations: directors of school district research and evaluation units. Most of the people we spoke to agreed that, while current federal regulations do not necessarily preclude use, they offer

little that helps to stimulate local evaluation use; their effect may, at best, be neutral. By implication, these district personnel were saying that, if it is to be a valuable partner in evaluation and its uses, the federal government must do more to promote evaluation use among the local schools and districts whose efforts it is supporting.

While directors agreed that the increased flexibility provided by Chapter 2 and (to a lesser extent) Chapter 1 is desirable, some felt that the new regulations were too nondirective. In this regard, Hastings' (1983) interviews with school district Chapter 2 administrators suggest that the feelings expressed by our respondents are felt elsewhere. Districts need some assistance, our respondents believed, in reconciling what is perceived as required for federal purposes and what is needed for local use. This concern is reflected in the following observations, especially those on historical precedent and on the perceived need to mount parallel evaluations.

Several directors were concerned about the force of historical precedent. Because of earlier state and district preferences (which grew out of earlier federal emphases), they were still required to administer a norm-referenced test for reporting purposes. But information produced by such tests was of limited use to them in carrying out their responsibilities. Further, all but one of the respondents agreed that these tests did not address teachers' interests or provide the kinds of information needed to monitor and adjust classroom instruction. Most directors felt that a good deal of their limited time and resources were given over to generating information which was state required in response to Chapter 1 regulations.

A few districts were trying to distill some locally-useful information from a norm-referenced test. But the directors felt that too much time was involved in having the tests scored and then returned for analysis and interpretation so that they would be of some use to curriculum specialists and classroom teachers. Consequently, these efforts detracted from the time they needed to generate information more specific to local instructional needs.

Several districts, in addition to administering and reporting the results of a norm-referenced test, used locally developed objectives-based measures and/or the tests accompanying curriculum materials. (See Burry, et al 1982 for a discussion of how widespread this practice is, as well as some of its implications.) While districts found the information from these additional tests instructionally useful, they asked why they had to conduct what was, in effect, a parallel evaluation. That is, for the reasons suggested above, they felt they should use a "respectable" norm-referenced test for external reporting purposes; given the limitations of the information provided by such tests, however, they felt at the same time that they had to resort to other devices for locally useful information.

The question raised by the directors was as follows: if such district-level efforts are necessary to produce formative data for district purposes, cannot these efforts also be used to satisfy external (summative) requirements? They asked this question even though they were aware that federal requirements imposed no particular kind of test; perhaps they (or their superiors) continue to believe that anything other than a standardized test is unacceptable.

This situation created a double dilemma. On the one hand, the research and evaluation directors realized that, to be locally useful, the evaluation should provide different types of information for different groups of users and that such an effort takes time: time to identify the needs and questions of various potential uses, time to develop or select appropriate tests or design other data collection procedures, time to win user support for the evaluation. On the other hand, the felt need to run a separate evaluation for external purposes also takes time, time that might be better spent addressing local questions and needs. Moreover, they felt that these questions and needs were precisely the ones in which the federal government should be interested but which were not reflected in the regulations.

Relative to the flexibility inherent in current federal requirements, directors mentioned that they had a hard time convincing their staff, resource specialists, and teachers to become involved in the design of the evaluation. Because of what they knew or believed about previous evaluations, personnel were hesitant about raising their own evaluation questions, reluctant to participate in the process of devising ways to answer them, and unwilling to believe that the evaluator would want to help them in the task of carrying out their day-to-day responsibilities. Thus, the current federal approaches to reducing external supervision and control -- far from stimulating staff members' interest in evaluation use -- seemed to arouse disbelief, anxiety, and hostility.

These observations seem to agree with the conclusions of our analyses of current federal requirements and their effects on local evaluations use.

Summary

In this paper we have drawn on our understanding of evaluation and its uses to examine current federal evaluation requirements.

It appears that, insofar as their language is concerned, Chapter 1 requirements are, as intended, potentially less burdensome than in the past. Chapter 2 regulations have even greater potential for increasing flexibility and decentralization. In both cases, the extent to which flexibility is increased depends, to some extent, on how states exercise their evaluation responsibilities. Some states may impose more rigid requirements than others, thus circumscribing the new freedom offered by federal regulations. Others may attempt to capitalize on their discretion but fail to stimulate local use. Therefore, the role of the states should be considered in terms of the contribution it makes to promoting the local uses of evaluation.

These considerations aside, however, the language of current Chapter 1 and 2 requirements is unlikely to stimulate local programs to capitalize on their new found flexibility to promote local evaluation uses. The same criticism also applies to the language of the requirements accompanying bilingual, handicapped, and vocational education programs.

Now, the language of federal evaluation requirements should probably not begin to suggest ways of building local uses into program evaluations. Suggestions, as we have seen, have a tendency to be viewed as hard and fast requirements, and such a tendency runs counter both to the federal desire to reduce unnecessary burdens at the local level and to the need for local districts to influence the use factors that are important in their particular setting. On the other hand, federal regulatory language should be

REFERENCES

- Alkin, M.C., Daillak, R., & White, P. Using evaluations: Does evaluation make a difference? Beverly Hills, CA: Sage Publications, 1979.
- Alkin, M.C., Jacobson, P., Burry, J. & White, P. Organizing for evaluation use: A handbook for administrators. Los Angeles: UCLA Center for the Study of Evaluation, 1983.
- Burry, J. Evaluation in bilingual education. Evaluation Comment, 1979, 6,(1), 1-14.
- Burry, J., Catterall, J., Choppin, B. & Dorr-Bremme, D. Testing in the nation's schools and districts: How much? What kinds? To what ends? At what costs? CSE Report No. 194. Los Angeles: UCLA Center for the Study of Evaluation, 1982.
- Council of Chief State School Officers Consolidation Evaluation Task Force. Memorandum: Final report and recommendations. Council of Chief State School Officers, Washington, D.C.: 1982.
- Daillak, R.H. A field study of evaluators at work. CSE Report No. 154. Los Angeles: UCLA Center for the Study of Evaluation, 1980.
- Gray, P.J., Caulley, D.N., & Smith, N.L. A study in contrasts: Effects of the Education Consolidation and Improvement Act of 1981 on SEA and LEA evaluation. Report No. 79. Portland, Oregon: Northwest Regional Educational Laboratory, 1982.
- Hastings, A.H. Snipping the strings: Local and state administrators discuss Chapter 2. Phi Delta Kappan, 1983, 65(3), 194-198.
- Stecher, B.M., Alkin, M.C., & Flesher, G. Patterns of information use in school level decision making. CSE Report No. 160. Los Angeles: UCLA Center for the Study of Evaluation, University of California, December 1981.