Utilization of Findings from Program Evaluations: Literature Review

Report

Prepared for:
Department of Health and Human Services
Assistant Secretary for Planning and Evaluation

Submitted by:
The Lewin Group, Inc.
Rachel Wright
Sarah Stout

January 16, 2008
Table of Contents

I. INTRODUCTION ................................................................................................................................. 1

II. DEFINING UTILIZATION .................................................................................................................... 2
   A. INSTRUMENTAL USE ......................................................................................................................... 3
      1. Characteristics of Instrumental Evaluations ............................................................................. 3
      2. Examples of Instrumental Evaluations ................................................................................... 4
   B. INFLUENTIAL USE ......................................................................................................................... 4
      1. Uses of Influential Evaluations ............................................................................................... 5
      2. Examples of Influential Evaluations .................................................................................... 5

III. FACTORS AFFECTING UTILIZATION .............................................................................................. 6
   A. EVALUATION QUALITY .................................................................................................................. 7
   B. INVOLVEMENT OF STAKEHOLDERS ........................................................................................... 7
   C. DISSEMINATION OF FINDINGS .................................................................................................... 9
   D. POLITICAL CLIMATE ................................................................................................................... 9
   E. AGENCY OVERSIGHT .................................................................................................................... 11
      1. Program Assessment Rating Tool ......................................................................................... 11

IV. CONCLUSIONS .................................................................................................................................. 12
   A. IMPLICATIONS FOR THE CURRENT STUDY ............................................................................. 13

REFERENCES ........................................................................................................................................... 15
I. Introduction

Evaluation, defined as “a social science activity directed at collecting, analyzing, interpreting, and communicating information about the workings and effectiveness of social programs,” is often commissioned to either obtain evidence that a program is providing desired results or to help illustrate how a program can be improved (Rossi et al., 2004). In addition to program accountability, program evaluation is also helps shape public policy. Specifically, Rossi et al. (2004) note that evaluations are conducted to:

- Satisfy accountability requirements of program sponsors,
- Improve program management and administration,
- Assess the value of new program initiatives,
- Help with decisions about whether or not a program should be continued, and
- Contribute to the universe of social science knowledge.

Program evaluations are essential to better existing social programs and create successful new social programs. However, studies of evaluations demonstrate inconsistencies in the extent to which findings are used to inform program and policy decisions. Thus, it is important to understand the extent to which evaluation findings are used and what factors have been associated with the successful utilization of findings. The Department of Health and Human Services Assistant Secretary for Planning and Evaluation contracted with The Lewin Group to examine how findings from recent program evaluations conducted by the Department of Health and Human Services (HHS) have been disseminated and used.

This study presents an exciting opportunity to explore how the findings of HHS-funded program evaluations are used and to relate that use to the key factors that likely determine how results are received and used. This literature review addresses the following key research questions in order to provide essential background information for the study:

- How is use of program evaluations defined?
- How have program evaluations been used both internally and externally to support or modify program policy, design, or operations?
- What factors facilitate the utilization of evaluation findings?
- How can the existing literature help guide the framework for the study of HHS evaluation utilization?

The remainder of this paper provides background information on each of these questions and reviews findings from previous research. While some of the literature provides evidence from
empirical studies on evaluation use, many of the findings cited in this document draw on the personal insights of researchers based on their experiences with evaluation use.¹

Findings from the literature on how and why evaluations are utilized will help inform the study of HHS evaluation use by providing a conceptual framework for determining how evaluations have been used and the factors that influence the use of evaluations. This paper is organized as follows:

- Section II reviews the various ways that evaluation findings can be used in order to help provide a definition of utilization.
- Section III highlights factors that affect evaluation utilization and discusses strategies for improving dissemination and use.
- Section IV provides a brief summary of this review and conclusions. This section also provides a conceptual framework for the study of HHS evaluations by highlighting types of use and examples, as well as factors that may help identify why an evaluation is used, or why it is not.

II. Defining Utilization

Evaluation utilization literature provides many definitions for what is meant by evaluation “use.” Findings may be used to make discrete decisions about a program, help educate decision-makers, or contribute to common-knowledge about a program, which may in turn influence later decisions about similar programs. Types of evaluation use can be divided into two broad categories: instrumental or direct use and influential or indirect use.

In the policy environment, evaluations may be used by a diverse array of stakeholders. Users include individuals and organizations formally linked to the program being evaluated, including program sponsors, program directors, and program practitioners. Evaluations also serve users with other relations to or stakes in the program (Hofstetter & Alkin, 2003). Information from evaluations may be used by:

- Managers of related programs may apply lessons learned to improve their programs,
- Administration officials or legislative policymakers responsible for program oversight, policy direction, and funding,
- Clients the program serves and advocacy groups that represent these clients,
- Social scientists who may use the evaluation results to build theory and improve their own research through the accumulation of knowledge, and

¹ When findings presented in this report are from empirical research, as opposed to researchers’ personal experience and insight, the review explicitly describes the study.
Utilization of Findings from Program Evaluations: Literature Review

- Others who may use evaluations to illustrate program performance and counteract public hostility and apathy toward social programs.

Another notion of an evaluation user is the “learning organization.” Weiss (1998a) explains that the notion of learning organizations is that “outside conditions are changing so rapidly that in order to survive, organizations have to keep learning to adapt.” Thus, organizations, as well as individuals, use evaluations to help modify organizational conditions so that programs may be improved. Evaluations can be used by both intended and unintended users in either direct or indirect ways.

A. Instrumental Use

Evaluation use may be most apparent when findings are used to make concrete decisions about a program. This “instrumental use” may include decisions about program funding, about the nature or operation of the program, or decisions associated with program management (Cousins and Leithwood, 1986).

1. Characteristics of Instrumental Evaluations

Instrumental use frequently occurs when the evaluator understands the program and issues, and does a good job conducting the study and communicating its results (Weiss, 1998a). Researchers report that their experiences with evaluations suggest that instrumental use is also common under the following conditions:

- Evaluation findings are non-controversial,
- Implied or recommended changes are easy to implement,
- The program environment is stable in terms of leadership, budget, types of clients, and public support, or
- A program crisis exists.

Other research suggests program size may affect whether findings are used instrumentally. According to Hofstetter and Alkin (2003), evaluation findings may be more visibly utilized in smaller-scale projects because the evaluators can develop personal relationships with key stakeholders in the evaluation. In smaller programs, there are also fewer actors in the political scene, and the context is less likely to be politically charged.

While program evaluations may be an important factor informing choices about programs, decision-makers pay attention to many other factors outside of evaluations of program effectiveness when making decisions about program management, operations and funding. These other factors include political support, the ease of implementing recommended changes, input from program participants and staff, and costs of changes.
2. **Examples of Instrumental Evaluations**

The Cash and Counseling Demonstration and Evaluation is an example of an instrumental evaluation, which was jointly funded by ASPE and the Robert Wood Johnson Foundation. Under Cash and Counseling, the Centers for Medicare and Medicaid Services approved Medicaid waivers for Arkansas, Florida and New Jersey to allow older and disabled Medicaid beneficiaries the flexibility to direct their own personal care services, including allowing them to pay family caregivers rather than traditional caregivers. The evaluation, conducted by Mathematica Policy Research, Inc., was rigorous (experimental). Participants were assigned randomly to the Cash and Counseling program or traditional agency services. Enrollment in the program began in 1998 and preliminary results from the evaluation were released in 2003, with a final report in 2007. Researchers found that participants in the program had fewer unmet care needs and higher satisfaction than those receiving traditional agency services, though concerns about the program included its higher costs (Brown, Carlson, et al., 2007).

The evaluation findings contributed to decisions by all three states to renew their waiver programs and to requests from an additional 11 states to implement Cash and Counseling programs (Kemper, 2007, Brown, Carlson et al., 2007). According to Kemper (2007), the policy influence of the evaluation can be attributed to several factors, including: rigorous evaluation design; wide evaluation scope, allowing researchers to analyze intended as well as unintended effects; the focus on implementation/operational issues, which yielded valuable lessons for other states implementing consumer-directed care options; and early dissemination of the results, which described the implementation experience and helped maintain interest in the demonstration programs.

In their study of the use of Drug Abuse Resistance Education (D.A.R.E.) evaluations, Weiss et al. (2005) found that the evaluation findings, as well as the influence of an overseeing body, were used by communities to make decisions about the implementation of drug education policy. Numerous evaluations of D.A.R.E. found that the program was not effective in preventing youth from using drugs. Interviews with school administrators and others working with the school drug prevention programs suggested that the evaluations were central to the decision to move away from the D.A.R.E. program.

Another example of instrumental evaluation is the California Greater Avenues for Independence (GAIN) welfare-to-work experiment. This project evaluated different approaches to the operation of GAIN in six county welfare programs in California. Riverside County emphasized quick entry into the labor market; other counties focused more on human capital development through education and training (Riccio et al., 1994). The evaluation estimated larger impacts of the program on employment and earning outcomes for welfare recipients in Riverside County, where there was a strong work-first focus, than in other counties that had a stronger training focus. The findings from the evaluation strongly contributed to California’s transformation of their welfare system into a more work-first oriented system.

B. **Influential Use**

Several studies of the use of evaluation findings suggest that while evaluations may have direct measurable effects on a program or policy, they may also be used more subtly to, for example, reduce uncertainty, increase awareness about a program and provide education, confirm
Utilization of Findings from Program Evaluations: Literature Review

existing viewpoints, and provide accountability (Hofstetter & Alkin, 2003). Evaluations, or an accumulation of multiple evaluations, can play an important role in changing conventional wisdom about policy problems and the potential impact of changes in policy (Kemper, 2003).

Evaluation influence has been defined as “the capacity or power of persons or things to produce effects on others by intangible or direct means” (Weiss et al., 2005). Evaluation influence can also have an impact on other institutions and events, beyond the program that is being evaluated. Michael Patton found that evaluations had been used by decision-makers in a variety of ways, including reducing uncertainty, increasing awareness, speeding up programmatic changes, and starting dialogues about the programs (Hofstetter and Alkin, 2003).

Also, “when evaluation adds to the accumulation of knowledge, it can contribute to large-scale shifts in thinking—and sometimes, ultimately, to shifts in action” (Weiss, 1998a).

1. Uses of Influential Evaluations

Influential evaluations may be used conceptually or educationally. Evaluations may influence common-knowledge about the most appropriate structure for a program, or help identify what program is ideal to address a particular problem (Cousins and Leithwood, 1986). Even if evaluation findings are not directly used, the ideas and generalizations from evaluations can be used by decision-makers to change the way policies and programs are thought about (Weiss, 1998a). Although it is not always easy to identify conceptual use of evaluation findings, it can become apparent when decision-makers use their new conceptual understanding to influence future policy.

Influential evaluations can be used symbolically to legitimize a policy or program. That is, evaluation findings may be used to justify a decision-maker’s preconceived notion about what program or policy is appropriate (Weiss et al., 2005). Evaluations can help foster support for policies “decided on the basis of intuition, professional experience, self-interest, organizational interest, a search for prestige, or any of the multiplicity of reasons that go into decisions about policy and practice” (Weiss et al., 2005).

The concept of influence recognizes that an evaluation can have either intended or unintended influences or a mix of the two, and that not all intended influences are explicit. An evaluation’s influence can occur at varying times during and after evaluation. As indicated by Karen Kirkhart (2005), the concept of influence is that it captures the impact of evaluations across the dimensions of intention and time. An evaluation influence can occur or become visible at the same time that the evaluation is occurring, at the completion of the evaluation through products of the evaluation and dissemination strategies, or the evaluation influence can be long-term, where effects of the evaluation may not be apparent for an extended period of time after the completion of the evaluation. Long-term influence may also occur when evaluation findings contribute to conceptual use where, as previously discussed, an evaluation—or accumulation of evaluations—adds to the continual process of dialogue and knowledge about a program over time.

2. Examples of Influential Evaluations

An early example of an influential evaluation was the Seattle-Denver Income Maintenance Experiment (SIME/DIME). This evaluation was conducted between 1970 and 1977 and a final
Utilization of Findings from Program Evaluations: Literature Review

A report was released in 1983. The evaluation tested the impact of various levels of income support on work effort by comparing an experimental group who received a cash transfer payment and SIME/DIME education and training services to a control group of families who only had access to the Aid to Families with Dependent Children (AFDC) program (the federal assistance program in effect from 1935 to 1996) (Brodkin & Kaufman, 1998). The evaluation found that there was a reduction in hours worked among the experimental SIME/DIME group, but that the reduction was relatively minor, about a 6 percent reduction for men and a 7 percent reduction for single-parent women (Brodkin & Kaufman, 1998). Furthermore, the reduction in hours worked was due largely to delays in entry into employment rather than a reduction in hours worked by those who were already employed.

Because of the small impact, both those who supported and those who opposed income supports attempted to use the evaluation as evidence to support their opinion. Proponents of cash transfer programs used the evaluation to promote their argument that this type of program would not cause a large retreat of workers from the labor force, while opponents used the evaluation to argue that the reduction in work hours (however small) confirmed worries that income supports undermine the work ethic (Brodkin & Kaufman, 1998).

Because the evaluation failed to reduce disagreements on this issue, some may say that the evaluation was not used. However, both opponents of the program and proponents used the evaluation to bolster their own arguments, and thus it was an influential evaluation because findings from the evaluation were used to support preexisting attitudes toward the program. This evaluation highlights the complexity in understanding the use of evaluations in the complicated political environment by illustrating the extent to which evaluation use can be subtle, and many not result in any significant policy changes, yet can still be influential in the policy debate.

III. Factors Affecting Utilization

While much of the research to date suggests that findings from a single evaluation may not frequently be used instrumentally to dramatically alter programs as a direct result of the evaluation, there is evidence to suggest that findings are used in many other ways that may have more indirect impacts on programs and approaches to policy. There are several factors that have been found to be associated with the likelihood that evaluation findings will be utilized in some way. The most widely cited factors affecting utilization include:

- Quality of the evaluation and the credibility of the evaluator,
- Involvement of stakeholders in the evaluation process,
- Dissemination strategies,
- Political climate at the time of the evaluation, and
- Oversight of the agency that manages the program.

A brief discussion of each of these factors follows.
A. Evaluation Quality

Quality evaluations are more likely to be used. Throughout the literature, there are many different definitions of evaluation quality, and many of the characteristics of quality that are used by researchers lack a clear operational definition.

In their meta-analysis of 65 studies of the use of evaluations conducted between 1971 and 1985, Cousins and Leithwood (1986) found that the most influential factors for evaluation utilization were quality, sophistication, and intensity of evaluation methods. Furthermore, in his study of evaluations, Henry (2003) found that influential evaluations were more likely to be of higher quality and more credible, as evidenced by being published in peer reviewed journals. Inaccurate evaluations, those with technical or methodological flaws, are more likely to be questioned and less likely to be used. Other researchers as well have also found the quality of the evaluation to be a key factor in determining the use of findings.

Because experimental evaluations are better able to demonstrate causality, it seems likely that findings from experimental evaluations might be used more than non-experimental evaluations. However, the literature that was examined for this review does not assess the extent to which experimental evaluations are used compared to non-experimental evaluations. Nonetheless, researchers have found that regardless of the type of evaluation conducted, high quality evaluations had several characteristics in common:

- Experienced evaluators with knowledge of the program area,
- Enough funding to conduct the evaluation,
- Sufficient period of time for the study, and
- Cooperative program personnel.

B. Involvement of Stakeholders

Stakeholder involvement in the evaluation matters. The level of participation by decision-makers, or stakeholders, in the evaluation process will affect the extent to which findings are used. Evaluation stakeholders are defined as anyone with a vested interest in the evaluation findings, and can include anyone who makes decisions or desires information about a program (Patton, 1997). Involving stakeholders may contribute to evaluation use because it helps tailor the study to the information needs of stakeholders, increase the credibility and validity of the evaluation from the perspective of the decision-makers, and help increase stakeholder knowledge about evaluations (Torres and Preskill, 1999). As Weiss (1998 a) notes that “the best way that we know to date of encouraging use of evaluation is through involving potential users in defining the study and helping to interpret results, and through reporting results to them regularly while the study is in progress.”

Including decision-makers in the evaluation process gives them a sense of increased confidence in the evaluation methodology, the quality of the information, and a sense of ownership of results (Shulha and Cousins, 1997). In their meta-analysis, Cousins and Leithwood found that several of the factors that influenced the use of evaluation were related to involvement of
stakeholders and stakeholder perceptions of the evaluation. Specifically, evaluations were most likely to be utilized if:

- Decision-makers found the study to be valid and important,
- Decision-makers found the evaluation to be relevant to their problems,
- The evaluation findings were consistent with decision-makers’ expectations,
- The evaluation maintains contact with decision-makers for an extended period of time after the end of the evaluation.

Henry (2003) found that influential evaluations had been responsive to a particular stakeholder or many stakeholders. Additional support for the impact of involving stakeholders in the evaluation process came from a survey of evaluators conducted in the 1970s. By conducting interviews with evaluators of 20 federal health evaluations, Patton and colleagues identified two key factors that affected the use of evaluations: “the personal factor” and political considerations (which will be discussed below). The “personal factor” was defined by Patton as “the presence of an identifiable individual or group of people who personally care about the evaluation and the findings it generates” (Patton, 1997). When an evaluation possessed this personal factor, or a “champion”, findings were more likely to be used, whereas when the personal factor was absent, findings were less likely to be used.

While stakeholder involvement may be critical to increasing the utilization of findings, it does raise some concerns. One is that stakeholders may lack the appropriate level of technical skill in evaluation methods and practices (Torres and Preskill, 1999). Another is that involvement of stakeholders in the evaluation may diminish the evaluator’s independence and objectivity. A considerable body of research suggests that responsiveness to stakeholders and the technical quality of the evaluation are likely to conflict (Bryk & Raudenbush, 1983; Greene, 1990; Patton, 1997; cited in Henry, 2003). Thus, while involving stakeholders has many benefits, care should be given to ensure that evaluators review ethical principles with stakeholders and clarify their own expectations about the involvement of stakeholders. While stakeholders can enhance the quality and usefulness of the evaluation by being involved in the design of the evaluation goals, purposes, and uses, many researchers suggest that stakeholders should not be involved in the technical aspects of the evaluation in order to maintain evaluator objectivity as well as to use the most sound methodological approaches as possible.

While involving stakeholders who are directly invested in the program can significantly increase the use of evaluations, neglecting to involve the entirety of stakeholders can lead to bias in the evaluation. The findings of evaluations are dependent upon what kinds of questions are being asked; if the program professionals are the only stakeholders involved in advising the evaluation, it may be biased by failing to examine questions that the program stakeholders fail to highlight. Weiss (1998a) notes that “studies that emerge from staff collaboration will not challenge the existing division of labor or lines of authority.” Thus, maintaining evaluator independence and objectivity may benefit from including a broad range of stakeholders rather than just those with direct program ties.
C. Dissemination of Findings

Compared to evaluation quality and stakeholder involvement in evaluations, less is known about the extent to which the method of disseminating evaluation findings influences their utilization. According to Lawrenz et al. (2007), to date, no data have been collected empirically about the effectiveness of various dissemination modes and strategies (e.g., publication in journals, presentations, Web, etc.) of program evaluation findings. A few studies have examined some components of effective dissemination strategies, including how widely the findings are disseminated, the timing of the dissemination, whether there are ongoing communications about the findings, the format in which the findings are presented, and whether there are mechanisms in place to track the implementation of findings. These factors are discussed briefly below:

- **Target Audience.** There is general agreement in the literature that broad dissemination of findings to the entirety of stakeholders is critical to their utilization (Shulha and Cousins, 1997; Weiss, 1998; Lawrenz et al., 2007; Preskill and Caracelli, 1997). Furthermore, according to Lawrenz et al., the dissemination strategies should be tailored for each audience in terms of scope, timing, and presentation format.

- **Timing.** As discussed further under Section D (Political Climate), the timeliness of the dissemination of evaluation findings can influence the extent to which they are used. If the results of a program evaluation are not made available within the window of time in which policymakers are establishing program priorities and making budget decisions about the program, the opportunity for the evaluation to substantially influence the direction of the program may be lost (Greenberg, Mandell, 1991). Several studies have also found increased use of evaluation findings when findings were communicated periodically rather than only once (Shulha and Cousins, 1997; Weiss, 1998; Lawrenz et al., 2007; Preskill and Caracelli, 1997).

- **Format.** While we did not identify studies that suggested that one method of disseminating findings was superior to another, at least one author discussed the importance of providing findings in a format that is “accessible” to the intended user (i.e. short and clear with charts and illustrations as well as executive summaries so that the reader can quickly absorb the key information).

- **Monitoring.** According to Dibella (1990), there needs to be an institutionalized method for following up on evaluation implementation recommendations for program improvement. For example, in addition to preparing reports and briefings, the status of implementing recommendations could be tracked, providing visibility to utilization—or lack of utilization (Dibella, 1990).

D. Political Climate

Political climate profoundly influences use of evaluation findings. In his study of 20 federal health evaluations, Michael Patton (1997) found that political considerations were one of the key factors affecting evaluation utilization. One way to think about the impact of the political climate on the use of evaluation is by differentiating between the impact of partisan politics and that of organizational politics.
Partisan positions of policymakers may influence how they perceive the findings of an evaluation. If the findings do not corroborate their intuitions or perceptions about the value or effectiveness of a program, they may be less likely to believe the evaluation findings and to use the findings for making future program or policy decisions (Greenberg and Mandell, 1991). Evaluations with findings opposed by an administration may be less likely to be disseminated widely and be used by the administration’s officials. Likewise the partisan politics of interest groups that are program stakeholders may influence the extent to which they use the findings of an evaluation in their advocacy efforts.

Recognizing that the political climate plays an important role in the utilization of findings, the Program Evaluation Standards, developed by the Joint Committee on Standards for Educational Evaluation, provided the following guidance for making evaluations politically viable: “the evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted” (The Joint Committee on Standards for Educational Evaluation, 1994).

Cousins and Leithwood found that organizational politics also play an important role in the use of findings. In their meta-analysis of evaluation utilization research, they found that evaluation findings were most likely to be utilized if the findings were not viewed as a threat by the program staff (Shulha and Cousins, 1997). In addition, findings were less likely to be utilized if key staff left the organization, there were internal debates or conflicts about the budget, rivalries existed between agencies, or if there was pressure placed on the evaluators by program operators and directors (Shulha and Cousins, 1997).

As discussed under Section C (Dissemination), the timing of the dissemination of evaluation findings also may influence the extent to which they are used. Using political scientist John Kingdom’s term, evaluations are most likely to be used when the policymakers have the findings when the “policy windows” are open. In other words, there are limited periods of time when key decisions about program priorities and budgets are made and if information about the evaluations does not reach decision-makers in time, the influence of the evaluation on program directions may be limited (Greenberg and Mandell, 1991). Depending on the program, these “policy windows” may be tied to the annual budgeting process, to the legislative reauthorization process, or to some other event. If the policy windows are tied to the annual budget, they may reopen every year, but if they are tied to reauthorization or another circumstance, it may be several years before they reopen. Also, if an evaluation is of an issue that is viewed as highly important, then the policy window can open multiple times, or may never close, and the evaluation can have multiple opportunities for influence.

Certain types of research are at a disadvantage in terms of timeliness. Social experiments, for example, have longer time lags than other types of research. As Greenberg and Mandel note, the Seattle-Denver Income Maintenance Experiment was launched in 1970, an interim report with early results was published in 1976 and the final report was not published until 1983. However, given the importance of the topic of this evaluation, the Negative Income Tax (NIT), the evaluation was influential, as discussed in Section II, even given the long amount of time it took to conduct the evaluation.
E. Agency Oversight

Another factor that may play a role in the extent to which findings are utilized is the level of dedication of the agency or entity that oversees the program to program improvement and evaluation. In the study of the use of D.A.R.E. evaluations referred to earlier, Weiss et al. (2005) found that the influence of an overseeing body was essential to communities’ decisions to incorporate the findings of the evaluations of the D.A.R.E. program into policy decisions. As mentioned, numerous evaluations of D.A.R.E. found that the program was not effective in preventing youth from using drugs and these findings were influential in decisions about whether to continue the program. However, another key feature of the decisions to discontinue the use of D.A.R.E. was a result of the mandates by the Safe and Drug Free Schools, which provides federal funding for school-based drug prevention programs (Weiss et al., 2005). While the mandates did not tell schools which program they had to use, they required that schools must use a drug prevention program that evaluations had found effective or promising. Safe and Drug Free Schools also provided the schools with a list of programs that had been found to be either exemplary or promising, and D.A.R.E. was not on the list at all. By mandating that schools use programs that had been found to be effective, Safe and Drug Free Schools played an important role in ensuring that findings from drug prevention program evaluations were put to use by schools.

In a study of five federal agencies (Administration for Children and Families, the Coast Guard, Housing and Urban Development, National Highway Safety Administration, and National Science Foundation), the GAO found that these agencies were able to successfully carry out and use evaluations by using a number of strategies that enhanced evaluation capacity (GAO, 2003). Agency strategies for building evaluation capacity included:

- Instituting an evaluation culture, or “a systematic, reinforcing process of self-examination and improvement,
- Building collaborative partnerships, for example with states, to obtain access to needed data and expertise for evaluations,
- Obtaining analytic expertise and providing technical expertise to program partners, and
- Assuring data quality by improving administrative systems and carrying out special data collections.

Cousins and Leithwood also found that evaluators are more likely to take advantage of factors that enhance utilization of findings when they are educated about the structure, culture, and politics of their program and policy communities (Shulha and Cousins, 1997).

1. Program Assessment Rating Tool

Agencies may increasingly have the capability to influence the extent to which evaluations are used given increased expectations for demonstration of program effectiveness. The Government Performance and Results Act of 1993 requires federal agencies to report their progress on achieving agency and program goals annually and established a foundation for
increasing government accountability. Building on the foundation laid by GPRA, the White House Office of Management and Budget (OMB) developed the Program Assessment Rating Tool (PART), which is also used to appraise program effectiveness by assessing the clarity of program design and strategic planning and by rating management and program performance (GAO, 2003). The goal of PART is to provide a consistent approach to assessing the performance of federal government programs to inform the federal budget development process.

In order to receive a high PART rating, agencies must demonstrate that they are collecting information about their performance through evaluations and other mechanisms and using that information for performance improvement and to justify requests for additional resources. Specifically, in the strategic planning section, PART asks “Are independent evaluations of sufficient scope and quality conducted on a regular basis or as needed to support program improvements and evaluate effectiveness and relevance to the problem, interest, or need?” Absence of results from evaluation is considered a deficiency.

In a 2003 assessment of the impact of PART on the federal budget formulation process, the GAO found that PART had provided a useful structure to OMB’s review of agency performance, but that challenges included limited information on program results and inconsistent use of that information among OMB staff. GAO concluded that PART may be most beneficial as a program management tool. In a follow-up report to Congress, GAO noted that among four agencies, including HHS, the PART process appeared to have resulted in greater emphasis on evaluation. Each of the agencies reported that an OMB recommendation through the PART process had prompted them to complete recommended program evaluations (GAO, 2005).

The President signed Executive Order 13450, November, 2007, “Improving Government Program Performance and guidance was sent to all Executive Agencies directing them to implement it. Each agency must designate a Performance Improvement Officer who sits on an executive branch Council; the officer is also responsible for overseeing, coordinating, and facilitating the development of the agency’s strategic and performance plans, program goals, and performance measures. During the coming year, agency staff expect that this new performance-related effort may influence the direction and content of evaluation funding decisions or activities.

IV. Conclusions

We want to understand the factors that influence use of findings because program evaluations are a key component for the betterment of existing social programs and the creation of successful new social programs. We also want to learn how to accurately gauge the extent to which evaluations are used to help improve or modify program policy, design, or operations.

Studies have found that individual evaluations may not frequently result in significant modifications to programs, but they are commonly used conceptually to add to common knowledge about a program area or influence decision-makers’ understanding of a program and what it is designed to do. Evaluations can also be used to legitimize a policy or program.
Finally, evaluations can influence policy or programs in a more indirect manner that can occur at various points in time after the evaluation has been completed.

Research suggests that there are several factors that influence the extent to which evaluation findings are used by decision-makers. The most important factors are evaluation quality, the involvement of stakeholders, broad dissemination of findings, and political considerations. The evaluator has control over three of these four factors, and thus can play a large role in increasing the likelihood that findings will be used. However, research also suggests that intermediary organizations and the government play an important role in promoting the use of evaluation through the broad dissemination of findings to a broad audience. In fact, just focusing on evaluations as an important component of the culture of the agency may increase the extent to which evaluations are used by placing more emphasis on their importance.

Furthermore, institutional mechanisms to help ensure program quality, such as the Government Performance and Results Act and PART, might encourage agencies to devote more attention to conducting and using evaluations if budget decisions take into consideration what agencies have learned from evaluations. Given this increased emphasis on performance improvement within the federal government, this study is both timely and relevant.

A. Implications for the Current Study

The intent of this document is to provide a general overview of the literature on the utilization of evaluation findings and a conceptual framework for the current study. While empirical research in this area is limited, we can draw on a number of findings and insights from the studies cited about the types of evaluation use and the factors that influence evaluation use and incorporate them into our study.

The findings of the literature review have particular relevance for how we design the remainder of the study. Specifically, the literature review raises important questions about how we gather information about the use of evaluations of HHS programs, including:

- Who should we query to best understand if and how the results of specific evaluations were used? Should we go beyond surveying program evaluation project officers and speak to other federal program officials and policy makers? Should we talk to contractors?

- Given that the impacts of evaluations may not emerge for a period of time, is it a limitation of the study that we are only looking back at evaluation reports completed during the four years ending September 2007? Can we use focus groups with other federal officials and/or the interviews with non-federal stakeholders to provide a longer-term perspective on evaluation use?

- Is there value in exploring a body of work in a particular program area as opposed to focusing only on individual evaluations?

- How best can we target our limited data collection from non-federal sources?
• Given that political climate may affect the use of evaluations, will focusing on a sample conducted under a single administration provide too limited a perspective? Should we use the non-federal stakeholder interviews to capture a longer-term perspective on the use of evaluations?

We plan to raise many of these questions and issues with ASPE, the HHS Stakeholders Committee, and the Technical Work Group. This literature review provides a context in which to have a conversation with these groups about the study design. It will also provide a context for how we interpret and frame the study findings.

In addition to helping with the overall design of the study, the literature review will be particularly helpful in designing the project officer survey. The purpose of the survey is to determine for a selection of HHS-funded program evaluations whether and how the findings were utilized. The findings of this literature review suggest that the survey design should recognize both instrumental (direct) and influential (indirect) uses by including questions that would capture both types of use.

The survey instrument will also include questions related to the project officers’ perceptions of the factors that influenced whether and how the evaluations of their programs were used. Based on our review of the relevant literature, we would suggest that we probe the project officers specifically about the impact of the following factors on evaluation use:

- Perceived quality of the evaluation,
- Role of various stakeholders at each stage of the evaluation process,
- Mandate to conduct the evaluation (e.g., OMB recommendation, legislative mandate),
- Political environment in which the evaluation took place,
- Methods and approach to dissemination,
- Agency/organization culture and support for the role of evaluation.

The framework described above of considering both direct and indirect uses of evaluation and these five key factors that appear to influence the use of evaluations may also serve to guide the organization of our briefings and final report.
References


Utilization of Findings from Program Evaluations: Literature Review


