

# The Mechanisms and Outcomes of Evaluation Influence

MELVIN M. MARK

*Pennsylvania State University, USA*

GARY T. HENRY

*Georgia State University, USA*

Past literature has identified several putative precursors of use, as well as alternative forms of use. However, important shortcomings still exist in previous work on use. In particular, inadequate attention has been given to the underlying processes that may mediate the effects of evaluation on attitude and action. In essence, a key part of the theory of change for evaluation itself is missing. To help fill this gap, we describe a framework designed to capture key mechanisms through which evaluation may have its effects. The framework includes change processes that have been validated in various social science literatures. It identifies three levels of analysis (individual, interpersonal and collective), each with four kinds of processes (general influence, attitudinal, motivational and behavioral). With a more comprehensive view of the mechanisms underlying evaluation's influence, the field can move forward in relation to its understanding and facilitation of evaluation's role in the service of social betterment.

**KEYWORDS:** evaluation outcomes; evaluation theory; evaluation use; influence; mechanisms

Evaluation is closely tied to the types of programs, policies and practices that affect people's lives, but is itself one step removed from the direct action of these endeavors; therefore most evaluators are drawn to the topic of use.<sup>1</sup> Use is the link between the day-to-day work of evaluation, on the one hand and those activities that could actually improve the lives of program participants and society, on the other. Concern about use has generated perhaps more empirical research on evaluation than any other topic (e.g. Alkin et al., 1979; Caplan, 1977; Cousins, 1996; Williams et al., 2002; Knorr, 1977; Patton et al., 1977; Preskill and Caracelli, 1997; Rog, 1985; Weiss and Bucuvalas, 1977) although empirical research on evaluation has been quite limited overall. In addition to research, considerable

theoretical and conceptual work on use has also occurred, leading to a variety of distinctions among types of use, such as instrumental use, conceptual use, symbolic use, process use and misuse (e.g. Cousins and Leithwood, 1986; Patton, 1997; Shula and Cousins, 1997; Weiss, 1977, 1979).

Originally, the evaluation literature was dominated by the idea of direct instrumental use: the notion that evaluation findings would lead to immediate and specific actions such as program continuation, expansion, revision or termination (Caracelli, 2000). Now, under the umbrella of use, evaluation theorists include a diverse range of possibilities, such as conceptual, symbolic and process use, as well as misuse, in addition to direct instrumental use (Knorr, 1977; Patton, 1997; Shula and Cousins, 1997; Weiss, 1977). Conceptual use, sometimes called enlightenment (Weiss, 1977), refers not to immediate decision making and action about the program or policy that was evaluated, but to more general learning that takes place as a result of evaluation. Symbolic use, an early concern in the literature, refers for example to the possibility that evaluation is used to justify a pre-existing position or simply to signify the purported rationality of an agency. More recently, Patton (1997) and others have raised awareness of process use, the most recent major addition to the catalog of types of use. Process use is distinguished from earlier types of use in that it is stimulated, not by the findings of an evaluation, but by participation in the process of evaluation.

Unfortunately, these categories of use are distinguished by qualitatively different attributes. Instrumental and conceptual use are each defined in terms of the *type of change* that occurs as a consequence of evaluation. For instrumental use, change occurs in actions, while for conceptual use, change takes place in thoughts and feelings. Symbolic use, in contrast, is defined not so much in terms of an outcome or effect of evaluation, but primarily by the *intent, real or perceived*, of an actor or organization, such as when a politician uses evaluation findings to justify a pre-existing position or when an agency is viewed as commissioning evaluation to appear rational, evidence-based and accountable. Process use, on the other hand, is defined by the *source* of the evaluation influence, referring to effects that are stimulated by the process of participating in an evaluation, rather than by evaluation findings. Presumably, however, participation in the evaluation process could contribute to changes in either attitudes or actions, and thus process use could overlap with either conceptual or instrumental use.

Whether because of or in spite of these different categories of use, the concept of use is central to the field of evaluation. This is perhaps best illustrated by the sense of purpose that use provides to working evaluators. Drawing on a survey of US evaluators, Preskill and Caracelli (1997) report that 99 percent of respondents agree that providing information for decision making and improving programs – the two classic forms of instrumental use – are motivations for doing an evaluation. In short, the link between evaluation and the betterment of social conditions is absolutely crucial as a collective *raison d'être* of evaluation. However, current conceptions of use are inadequate in fully establishing this link, in part because they lack a clear moral anchor (Henry, 2000) and in part because of their inadequate attention to the underlying change processes that mediate evaluation's influence.

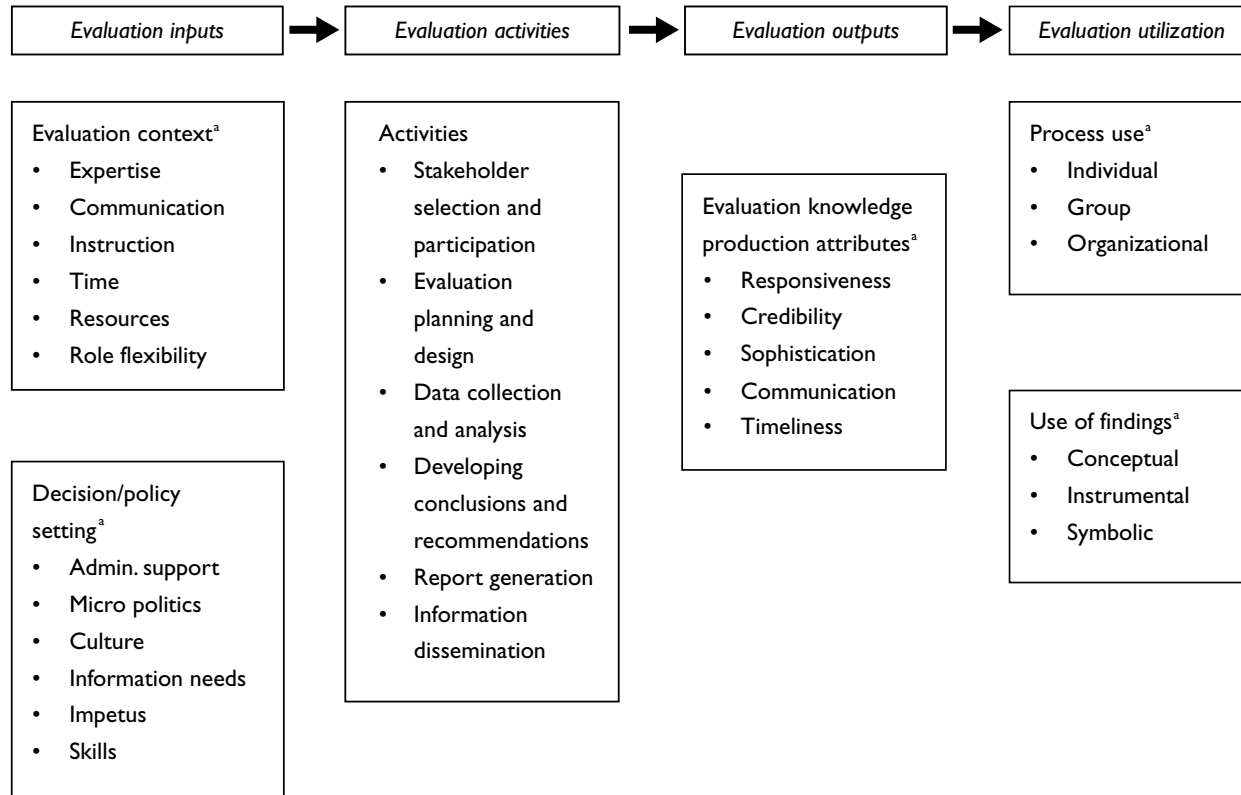
## **Current Theories of Use**

An analogy can be drawn between current theories (or models) of use and the logic models or program theories that many evaluators apply to the programs they are evaluating. Past writings about evaluation, in a sense, sketch out the model that many evaluators have implicitly adopted as to how evaluation can (and perhaps should) lead to use. In a recent review of use in the context of participatory evaluation, Cousins (2003) makes the applicability of logic models to evaluation use more explicit than usual, offering a model of utilization that looks strikingly like a program logic model. He incorporates two categories of evaluation utilization, process use and use of findings. Process use, in Cousins' model, can occur at three levels, individual, group or organizational. Use of findings also has three categories, the now familiar conceptual, instrumental and symbolic.

In Figure 1, we selectively adapt the conceptual framework of Cousins (2003), making more explicit the analogy to logic models and extending the model to include evaluation that is not explicitly participatory. In Figure 1, evaluation practice is shown as influenced by inputs, including the evaluation context and the setting within which the evaluation is being conducted. 'Evaluation context' involves the human and other resources allocated to the evaluation, while the 'decision/policy setting' includes the cultural, political and informational aspects of the organization(s) involved in implementing the program or policy and in initiating and sponsoring the evaluation. These inputs lead to evaluation activities such as stakeholder involvement, data collection and analysis, and dissemination. In turn, these evaluation activities generate the outputs of the evaluation. For the output stage, we draw from Cousins (2003), who refers to evaluation knowledge production.<sup>2</sup> Attributes of the knowledge produced by the evaluation are expected to influence the utilization of the evaluation, whether process use or use of the findings, or both.

Although useful in the same way that logic models are useful for program planning, contemporary theories of use (or evaluation utilization) are simultaneously impoverished and overgrown. Conceptions of use are impoverished in that, while existing models describe predictors of use and types of use, there is a critical missing link. Current models of use are generally silent on the range of *underlying mechanisms* through which evaluation may have its effects. Identifying these mechanisms is not merely a theoretical exercise. Just as attention to underlying mechanisms can be important in understanding, judging and improving programs and policies (e.g. Mark et al., 2000; Pawson and Tilley, 1997), a focus on underlying processes of evaluation influence is critical not only for evaluation theory and research, but also for guiding practice to better influence attitudes and action. We respond here to this impoverishment in the use literature, by outlining and conceptually classifying a set of mediators that plausibly may underlie the effects of evaluation on attitude and action.

While impoverished, in the sense of not attending adequately to underlying processes, current models of use are in other senses overgrown. The seemingly simple concept of 'use' has taken on alternative, partially overlapping, but



<sup>a</sup> Selected elements drawn from Cousins (2003)

Figure 1. Evaluation Logic Model

conflicting meanings. In part, the concept of use is overgrown in the sense that the multiple forms of use (instrumental, conceptual, symbolic and process) overlap; a major example is that process use can stimulate either action or understanding, that is, either instrumental or conceptual use. The problem with 'use' also includes the ambiguity of some key constructs and the absence of a history of developing rigorous indicators; presumably, for example, most evaluators believe they know what conceptual use or enlightenment is, but how should it be measured? Also, the problem in part is that 'use' is sometimes treated simply in a *descriptive* way, referring to whether or not some potential consequence of evaluation (such as an instrumental decision) *did* happen; while at other times 'use' is applied as a *normative* concept or guiding purpose for evaluation, referring to consequences of evaluation that *should* happen. When treated normatively (e.g. 'find an intended user'), attention to use can push evaluation in directions that may be undesirable from the vantage of other stakeholders (Henry, 2000; Williams et al., 2002). In addition, 'use' is an overgrown thicket partially because very different positions have been advocated as to the *scope* of evaluation use. Some evaluators restrict evaluation use (e.g. Alkin, 2003; Hofstetter and Alkin, 2003) to the specific environment and general time frame in which the evaluation was conducted, while others treat the domain of evaluation use more broadly, including changes that occur outside the original environment.<sup>3</sup>

As a result of these various forms of overgrowth in the concept of 'use', evaluators may not have a common understanding of what it means for an evaluation to be used, or of what another evaluator means when she refers to use. One response to such overgrowth within the taxonomies of use is to rely instead on an alternative concept, evaluation influence (Kirkhart, 2000). But this is not enough, for the real risk is that we would move from one thicket to another. Thus, it is also essential to work to develop a more detailed and more specific framework and terminology of influence as a way to clear out some of the conceptual thicket that has grown around the concept of use. In response to this need, in the remainder of this article we briefly present a framework and model intended to help alleviate the concurrent problems of impoverishment and overgrowth in the literature on use. In the next section, we present and classify several processes that may underlie evaluation's influence on attitudes and action. Greater attention to underlying processes should enable evaluators to better study, understand, facilitate and communicate about evaluation influence. We then describe several elements of a theory of evaluation influence, by expanding on Figure 1. This alternative framework posits specific outcomes and indicators that can be useful for research on evaluation influences, for planning aimed at achieving evaluation influence, and for avoiding the current conceptual and terminological thicket of use.

## **The Mechanisms Underlying Evaluation Influence: A Classificatory Model**

After describing the shopworn character of the term ‘use’, Kirkhart (2000) argues instead for a focus on *evaluation influence*. Unlike the narrower view of evaluation use, evaluation influence explicitly includes both changes that take place at the location and general time frame of the evaluation and changes that take place elsewhere and later. Henry and Mark (2003) further suggest that the focus of a theory of evaluation influence should be on those outcomes of evaluation, such as changes in attitudes about service delivery practices or changes in policy, that plausibly lead toward (or away from) the ultimate goal of social betterment. The framework presented in Table 1 below indicates the kinds of changes we see as falling within the scope of evaluation influence. In addition to arguing for a shift from use to influence, Kirkhart describes three dimensions related to evaluation influence: source, time and intentions.

### ***A Framework of Mechanisms***

Omitted from Kirkhart’s scheme, and missing from the evaluation use literature in general, is a detailed listing of the *mechanisms* through which evaluation may achieve influence and of the specific outcomes which would indicate that influence had occurred. Henry and Mark (2003) provide a research-based list of underlying mechanisms through which evaluation findings and processes influence attitudes and actions. These mechanisms are drawn from several areas of the social and behavioral sciences, and in most cases have been subject to considerable empirical investigation and validation as important change processes. Henry and Mark (2003) classify these mechanisms, which also include potential consequences of evaluation, in terms of three ‘levels of analysis’: individual, interpersonal and collective. Here we extend the work of Henry and Mark (2003) by expanding their list of mechanisms and by developing a more systematic, two-dimensional framework for organizing the alternative mechanisms that may mediate evaluation influence. In addition to the three levels of analysis (individual, interpersonal and collective), here we classify mechanisms into four kinds:

- general influence processes;
- cognitive and affective (or attitudinal) processes;
- motivational processes; and
- behavioral processes.

The result is the classification scheme shown in Table 1.

*General influence processes*, shown in the top row of Table 1, are the fundamental architecture of change.<sup>4</sup> They are likely to set into motion some change in the cognitive/affective, motivational or behavioral processes. Take one example from Table 1, at the individual level: elaboration can lead to changes in attitude valence (positive or negative) or in behavior. Elaboration refers to a cognitive process whereby an individual thinks systematically, or as Petty and Cacioppo (1986) put it, engages in ‘central processing’ about an issue based on information (such as an evaluation report). Elaboration can be measured by

*Mark and Henry: The Mechanisms and Outcomes of Evaluation Influence*

Table 1. A Model of Alternative Mechanisms that May Mediate Evaluation Influence

Type of Process/Outcome	Level of Analysis		
	Individual	Interpersonal	Collective
General influence	Elaboration Heuristics Priming Skill acquisition	Justification Persuasion Change agent Minority-opinion influence	Ritualism Legislative hearings Coalition formation Drafting legislation Standard setting Policy consideration
Cognitive and affective	Salience Opinion/attitude valence	Local descriptive norms	Agenda setting Policy-oriented learning
Motivational	Personal goals and aspirations	Injunctive norms Social reward Exchange	Structural incentives Market forces
Behavioral	New skill performance Individual change in practice	Collaborative change in practice	Program continuation, cessation, or change Policy change Diffusion

assessing how much time or effort an individual spends thinking in response to a message. An evaluation report, a conversation about an evaluation, or a newspaper article about an evaluation could trigger such cognitive processing. For example, a recently publicized evaluation about the positive effects of a preschool program may cause a reader at another location to think more about her views on educating children at early ages. Such a change may be exactly what some evaluators consider enlightenment. Of course, an evaluator would be interested not only in whether someone engaged in elaboration, but also in what if any changes this led to in the person's attitudes, motivations and actions. Still, elaboration itself is an important immediate consequence of evaluation, which might in turn produce a change in the individual's opinion about early education programs and, perhaps, subsequent change in behavior.

General influence processes can occur at all three levels, the individual, the interpersonal, and the collective, as indicated in Table 1. Consideration of these influence processes is important for understanding how evaluation can influence attitudes and actions. At the same time, the general influence processes are not likely *by themselves* to be important vis-a-vis social betterment; rather, their importance can be judged by whether or not they lead to other intermediate and long-term outcomes. That is, the general influence processes do not in isolation reveal whether evaluation has helped lead to (or away from) social betterment. For example, to know that a reader elaborated on the findings of a preschool program evaluation does not in itself tell you if any meaningful and important

change occurred. General influence processes are primarily of interest because they may (or may not) help stimulate the outcomes of greater practical interest, that is, the changes in beliefs and feelings, motivations and actions.<sup>5</sup>

The second, third and fourth rows of Table 1 contain, respectively, the cognitive and affective (or attitudinal) states, the motivations, and the behaviors that may result from evaluation's influence. *Cognitive and affective processes* refer to shifts in thoughts and feelings, such as attitude valence. As an example, the dated, early model of evaluation use seemed to assume the existence of central decision makers who would change policy or fund programs, if only evaluation would alter their opinions. More recent discussion of conceptual use also highlights the importance of cognitive and affective consequences of evaluation. *Motivational processes* refer to goals and aspirations, to human responses to perceived rewards and punishments. For instance, recent efforts to affect classroom teaching by setting standards (e.g. high-stakes testing) assume in part that behavioral changes among individual teachers can be influenced by standards that change the teacher's goals and aspirations.<sup>6</sup> Motivational processes have received less attention than attitudinal or behavioral processes in the traditional literature on use. However, motivational processes may be important in influencing practitioner behavior – and thus may be more important as intermediate than as long-term outcomes. *Behavioral processes* refer to changes in actions; these would include changes in a teacher's instructional practices at the individual level or a government policy change at the collective level. Thus, behavioral processes often comprise the long-term outcomes of interest in a chain of influence processes; nevertheless, behavioral processes can also appear as intermediate outcomes in a longer influence chain. For instance, evaluation findings may contribute to a policy change (e.g. implementation of standards), and that action may be intended to affect individual practitioners' motivation and thus their practices.

The entries in the individual cells of Table 1 list specific processes of each kind, at each level of analysis, drawn from various empirical literatures. Space limitations here preclude us from detailing and illustrating each of the specific entries in the cells of Table 1, though we describe several of them below in discussing the framework and its value. (For descriptions and brief illustrations of most of the entries in Table 1, see Henry and Mark [2003]; all are also discussed in detail elsewhere in the social sciences.)

Here we take a broader perspective, focusing on the conceptual framework represented in the column and row headings of the table and on the role of underlying mechanisms in a broader theory of evaluation influence. One benefit of attending to such mechanisms is that evaluators can better capture (and plan for) the multiple processes that lead from one to another, cascading forth to mediate the consequences of an evaluation. In other words, evaluation may trigger more than one of the processes in Table 1, with one process stimulating another. To take a relatively simple example, evaluation findings might lead to:

1. an individual's elaboration of her thoughts about a program, which might
2. make her attitude toward the program more positive in valence, which might in turn



*Mark and Henry: The Mechanisms and Outcomes of Evaluation Influence*

3. lead her to take on the interpersonal role of a change agent within her organization, which might
4. result eventually in reconsideration of organizational policy and ultimately in
5. collective-level policy change.

We return to the notion of multiple, linked influence processes below, with a case example. As even the preceding simplified example illustrates, each of the entries in Table 1 can play at least two key roles. Each can be an *outcome* of evaluation; each can also be an *underlying mechanism*, leading in turn to some other outcome from Table 1. As such, each individual process can be a short-term, intermediate or long-term evaluation outcome in the often-complex pathways to social betterment. Because the elements in Table 1 can play the dual roles of an outcome of evaluation and a mechanism that stimulates other outcomes, we often refer to them as ‘processes’.

Returning to Table 1, within each row there are corresponding processes as you move across the different levels of analysis. This is sensible, given that the rows are defined in terms of different kinds of processes. For example, among the cognitive and affective processes, agenda setting at the collective level corresponds to salience at the individual level, and a local descriptive norm is roughly analogous to these two at the interpersonal level.<sup>7</sup> Similarly, in terms of behavioral processes, there is a fairly clear correspondence as you move from the individual, interpersonal and collective levels of analysis, respectively, between individual change in practice, collaborative change in practice, and policy change, respectively. Such correspondences can also be found within the general influence processes: elaboration at the individual level, persuasion (and the dialogue that often is involved in persuasion efforts) at the interpersonal level, and the various forms of collective deliberation, are roughly corresponding forms of information processing at different levels of analysis. These correspondences suggest that, for some purposes, it may suffice to use the general framework of the three levels of analysis and the four kinds of mechanisms/outcomes. For other purposes, however, it will be more helpful to think about specific processes within each cell of Table 1.

***Traditional Forms of Use in Terms of the New Framework***

How do the familiar forms of use translate into the new framework? For the two most often discussed forms of use, that is, instrumental use and conceptual use (or enlightenment), Table 1 provides a more specific way of describing and operationalizing them. In general, instances of instrumental use fall within the behavioral row of Table 1. Although early discussions of instrumental use focused on the kinds of collective actions in the right-hand column of the table (e.g. Weiss, 1977), more recent work includes attention to behavior changes in the individual practitioner (e.g. King, 2002) and the collaborative group (e.g. Preskill, 1994). By disentangling the kinds of actions that correspond to the different levels of analysis, Table 1 provides a clearer picture of the potential effect of evaluation on behavior, in comparison with the classic but conceptually overgrown concept

### *Evaluation 10(1)*

of instrumental use. And such clarification is important: for example, if textbooks tell evaluators only how to try to achieve instrumental use, this ignores the fact that the pathways to change in individual practice are likely to differ greatly from the pathways to collective change including policy change (Weiss, 1998).

Conceptual use corresponds to the cognitive and affective processes row of Table 1. In our reading, discussions of conceptual use are often ambiguous or inconsistent regarding whether enlightenment occurs at an individual or collective level, or both. In contrast, the framework in Table 1 makes it easy to differentiate cognitive and affective effects at different levels of analysis. Moreover, with the processes listed in the table, it is far easier to be precise about what specific kind of change has occurred. For instance, at the individual level, Table 1 reminds us that evaluation might not change an individual's opinion, positive or negative, about a program, but might instead make some aspect of it more salient. An example from the US occurred during recent evaluations of welfare reform. These evaluations made children's well-being more salient as a possible effect of reform, in the minds of important evaluation audiences (Henry and Mark, 2003). This is a potentially important form of evaluation influence, but the broader notion of conceptual use might not lead evaluators to think of salience as a specific kind of outcome, distinct from traditional opinion change. Again, the distinctions matter. In research on evaluation's effects, for example, different kinds of measures are needed to assess a change in salience than to measure opinion change.

In short, then, instrumental use corresponds to the behavioral row of Table 1 and conceptual use to the cognitive and affective row. In contrast, symbolic use refers to a select set of interpersonal or collective influence processes. Specifically, symbolic use includes justification at the interpersonal level and 'ritualism' at the collective level (note that ritualism – the ritualistic use of evaluation to signify rational decision making – is not likely to trigger other forms of evaluation influence). It should perhaps not be surprising that symbolic use has a more limited place in the framework of Table 1 than does instrumental or conceptual use: it is a narrower category of use, and accordingly has received considerably less attention in the literature.

In contrast to the other traditional forms of use, process use does not correspond to specific rows or processes of Table 1. Instead, as noted previously, process use means that some change has arisen because of the process of an evaluation rather than because of an evaluation finding. Accordingly, process use, in our view, cannot be translated into the mechanisms of Table 1. Instead, process use is defined by whether influence is triggered, in the language of logic models, by evaluation activities rather than by evaluation outputs (i.e. findings).

In sum, the framework in Table 1 adds clarity relative to the commonly used, broad and overlapping categories of use. Table 1 can also be employed to highlight differences across alternative spheres of evaluation practice. For example, the right-hand column of Table 1 will be very important for those interested in the effects of evaluation on policy within governments (though the other columns will also be important in the often complex chains that may *lead to* policy effects). In contrast, the individual and interpersonal columns of Table 1 will generally be

the locale of action for those interested in the influence of evaluation on practitioner behavior, such as teacher practices (see Weiss [1998: 264–5] on the distinction between use of evaluation for policy and for practice).<sup>8</sup>

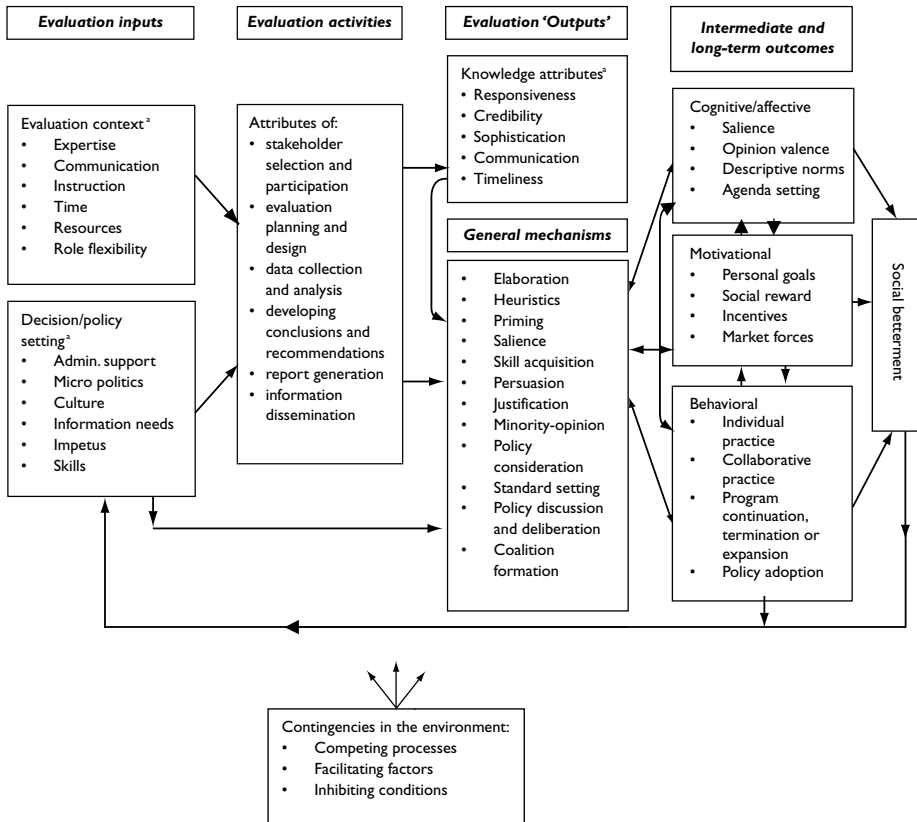
The framework in Table 1 thus can eliminate the confusion that may arise when evaluators use the same term (e.g. conceptual use) to refer to different change processes at different levels of analysis. In the next section, we move from categorizing evaluation influence processes, and begin to embed these processes in a more comprehensive theory of evaluation influence.

## **Toward a Comprehensive Theory of Evaluation Influence**

Figure 2 provides a revision of Figure 1, integrating into that earlier figure the processes from Table 1. For the sake of simplicity, we have not repeated from Table 1 the distinction among the three levels of analysis in the already-complex Figure 2. Simply for reasons of space, we also have not included *all* of the processes from Table 1 in the figure. In addition, we have listed the general influence processes separately from the cognitive and affective, motivational and behavioral processes of Table 1, and we have labeled the latter set of processes as intermediate and long-term outcomes.

Figure 2 illustrates several key differences between extant conceptions of use and the kind of new theory of evaluation influence we are suggesting. First and foremost, Figure 2 includes our set of underlying mechanisms. These influence processes, taken from Table 1, are expected to mediate the effects of evaluation activities on evaluation outcomes. The framework summarized in Figure 2 thus offers benefits of greater theoretical comprehensiveness. The change mechanisms, absent in Figure 1 but included in Figure 2, can fill a major gap that has existed in what might be called the program theory or theory of change for evaluation itself. That is, the past literature has identified possible precursors of use (e.g. credibility of the evaluation, stakeholder involvement), as well as general forms of possible use (e.g. instrumental, conceptual), but for the most part has ignored the change processes through which evaluation influences attitudes, motivations and action.

A second, very general benefit of the model in Figure 2 is that it can help organize a number of different hypotheses that could guide future research on evaluation influence, and also help guide practitioners' efforts to increase influence. Some hypotheses, of course, involve the direct role of the underlying processes presented in Table 1 (e.g. when does evaluation influence result from a change in salience, rather than traditional opinion change?). Several other hypotheses involve the way that the older parts of the model (the ones presented in Figure 1) exert their influence. Consider as an example the 'evaluation knowledge production attributes'. Figure 2 shows only one linkage from these attributes, directly to the evaluation mechanisms. This reflects our expectation that, if factors such as responsiveness and credibility have any effects, it is because they stimulate one or more of the mechanisms in Box 1. For example, if responsiveness has any effects, it may be because a more responsive report increases evaluation consumers' motivation to engage in elaboration (Petty and Cacioppo, 1986).



<sup>3</sup>Selected elements from Cousins (2003). Categories in bold taken from Table 1; see table for complete listing.

Figure 2. Schematic Theory of Evaluation Influence

As another example, consider the model's treatment of evaluation activities. Williams et al. (2002) have recently described varying practices in the EU with respect to evaluation activities such as dissemination. Our model suggests that, if these and other evaluation activities make a difference, it may be because of their direct effect on the evaluation mechanisms and processes (e.g. more frequent reporting may increase elaboration). Alternatively, evaluation activities can exert effects indirectly, by impacting 'evaluation knowledge production attributes' (e.g. good dissemination practices may increase perceived responsiveness). As a final example, the model posits a slightly more complex set of paths through which attributes of the 'decision/policy setting' may make a difference. Recent work has highlighted several attributes of the decision/policy context, including the location of the evaluation function in the larger organization (e.g. in an audit or planning unit vs a dedicated evaluation unit). In other recent work, Walter et al. (2003a) discuss the role of 'pull' factors:

*Mark and Henry: The Mechanisms and Outcomes of Evaluation Influence*

the demand for evaluative information that may exist in the policy setting. As shown in Figure 2, we hypothesize that these factors and other attributes of the decision/policy setting can influence evaluation outcomes, not only by any effects they have on the underlying mechanisms and processes (e.g. evaluators who demand evaluation information may be more likely to elaborate on it), but also by virtue of their effects on evaluation activities (and any subsequent effects of these evaluation activities).

Box 1 summarizes several benefits of the model presented in Figure 2.

***Influence Pathways***

Another noteworthy feature of Figure 2 is the set of linkages between the three different levels of evaluation outcomes and the bidirectional pathways between the underlying mechanisms and the outcomes. (These pathways do not imply simultaneous bidirectional causation, but rather the potential for iterating sequences of mechanisms and outcomes.) These potential linkages are included because complex chains of influence can exist. For example, an individual change process may occur first, stimulating an individual-level outcome that then leads to an interpersonal-level change process, and so on. These potentially complex linkages can be described, either graphically or narratively, by listing specific processes from Table 1 (and, preferably, indicators of them), in sequence, to

*Box 1. Selected Benefits of the New Framework of Evaluation Influence*

- Includes specific influence processes in the model (see Table 1 and related text for more detail)
- Draws on research traditions outside of evaluation, which can help guide development of indicators and generation of hypotheses (see Henry and Mark [2003] for more detail)
- Clarifies the nature of evaluation influence in terms of kinds of outcomes and level of analysis (see Discussion section, as well as Table 1 for more detail)
- Makes concrete predictions about the general relations between different components of the logic model of evaluation (see Figure 2 and text for more detail)
- Replaces previous notions of instrumental, conceptual, symbolic, and process use with more specific (and less overlapping) representations (see previous section for more detail)
- Explicitly includes contingencies in the environment, both facilitating and inhibiting factors and other processes that compete with evaluation influence processes (e.g. advocacy groups who may argue positions contrary to evaluation findings) (see Discussion section for more detail)
- Explicitly includes social betterment, the ultimate desired outcome by which evaluation's consequences ultimately should be judged (see Mark et al. [2000] regarding the use of social betterment as a criterion for evaluating evaluation)
- As a result of all of the above, holds the potential to better guide research and further theoretical development on evaluation's influence, as well as guide the development of better influence plans for evaluation practice (see Discussion section for further detail)
- Also may provide the basis for more thoughtful discussion about the responsibilities of the evaluator for evaluation influence (see Discussion section for more detail)

describe the *pathways* through which evaluation achieves its eventual influence (Henry and Mark, 2003). Writings that emphasize the study of underlying mechanisms in evaluation (e.g. Pawson and Tilley, 1997) might be interpreted as suggesting that there is only one relevant mechanism in a given case. To the contrary, the pathways to evaluation influence will often be complex and even circuitous, involving the concatenation of multiple mechanisms. As the following example illustrates, the processes of Table 1 can serve as linked steps in a cascading sequence of processes that leads to the ultimate effects of evaluation on attitudes, motivations and actions.

### ***A Case Example: The Influence of Performance Monitoring in US States***

In the US, 47 out of 50 states have enacted requirements to monitor the performance of all public agencies, program by program (Melkers and Willoughby, 1998). These performance-monitoring systems are now a major investment in state-level evaluation. In large part, these systems have been created so that performance information can influence each state's final appropriations bill, which sets the funding for state agencies. To gauge the influence of performance monitoring, Melkers and Willoughby (2001; Willoughby and Melkers, 2001) surveyed two major sets of respondents in each state. One consisted of budget officers in the executive branch of government, from agencies and from the state's central (executive) budget office. The second group of respondents were budget officers in the legislative branch of government. Although more distant from the agencies (where initial budget proposals begin), the legislative budget officers are involved in the legislature's work on the appropriations bill.

In early studies of evaluation use, researchers often asked key players in major policy or program decisions about their use of information from evaluations. If Melkers and Willoughby had adopted this approach, they might have interviewed only the legislative budget officers, who were directly involved in the legislative process. In fact, the legislative budget officers reported limited influence. For example, they rated the influence of performance measures on final appropriations decisions about midway between 'not important' and 'somewhat important'. This and other findings at the legislative level would seem to suggest that performance measures were not very influential. However, other evidence from Melkers and Willoughby allows us to trace a more complex evaluation pathway, and also to see more influence than would have been apparent if the (legislative) end-users alone had been interviewed.

The pathway begins with the performance-measurement system, an evaluative system:

1. making outcomes in general *more salient* among budget officers in the agencies,<sup>9</sup> which
2. consequently made outcomes more important in the agency-level budget development process (it is at this level where key initial decisions are made about program expansion and contraction). Subsequently,

*Mark and Henry: The Mechanisms and Outcomes of Evaluation Influence*

3. findings from the performance measurement system were used to help *justify* the agency's proposals to the central executive budget office. Next,
4. outcome-based arguments, grounded in the performance measures, were used *persuasively* in the centralized executive-level deliberations about appropriations.
5. The influence of the performance measures was then embodied within the *draft appropriations bill* proposed by the state's chief executive to the legislature. At that point,
6. the legislative budget officers, from their vantage point, could not see all the direct influence of the performance measures, because the earlier influence was 'built into' the draft legislation.

In contrast, among executive branch actors, considerably more influence was visible. Melkers and Willoughby (2001) and Willoughby and Melkers (2001) report findings for items related to the influence processes in the executive branch (Steps 1–5 above). Across these steps in the influence pathway, executive budget officers' ratings of the influence of the performance measures generally are about midway between 'somewhat important' and 'important'.

This case can also be examined in terms of the levels of analysis from Table 1. It appears the underlying processes started at the individual level (i.e. increased salience among individual budget developers), moved to the interpersonal level (i.e. both justification and persuasion), and finally moved to the collective level (i.e. the drafting of appropriation legislation and final budget adoption). The pathway we have identified based on Melkers and Willoughby's research is, we believe, more revealing and more persuasive than simply stating that some level of end-use did or did not occur. Nevertheless, other steps may also have existed in this case, including additional shifts across levels of analysis, but these were not the focus of Melkers and Willoughby's investigation. Their findings do however also make it possible to conclude that some mechanisms did *not* operate. In particular, there is no indication of ritualism, with 75 percent of all the budget officers reporting that the performance measures were of no importance as a means to appease the public.

This case example highlights four points. First, if the underlying mechanisms and intermediate outcomes were ignored and accounts were obtained only from end-users about the collective action, the results would have been seriously biased in the direction of finding little or no evaluation influence. Second, detailing the influence pathway not only reduces this negative bias. In addition, with the more complete and coherent story a pathway provides about how influence unfolded, the claim that 'evaluation influence has occurred' becomes more credible. The reason for this claim should be clear to those familiar with various evaluation approaches that use program theory to strengthen causal claims (e.g. Mark et al., 2000; Pawson and Tilley, 1997). Third, the framework offered in this article appears to be capable of capturing the influence pathways emanating from evaluation process and findings. At least in the case of Melkers and Willoughby's study of the effects of performance measurement, we were readily able to translate documented influence processes into a coherent influence pathway using the



elements of Table 1. Finally, as Carol Weiss (1998) noted long ago, evaluation findings are only one of the considerations that affect policy and program decisions. In this case, the influence of evaluation appears to have been moderate but important.

## **Discussion**

Before turning to the broad implications of the framework presented in Table 1 and the model presented in Figure 2, we offer some caveats.

### ***Limitations***

Although the framework offered in this article has important potential benefits, we certainly do not believe Table 1 provides the ultimate, final classification and listing of the mechanisms that underlie evaluation influence. The structure and individual entries of Table 1 are intended as a useful starting point for future theoretical and empirical work, rather than as a final product. For example, given the growing attention in the literature to the intersection of evaluation and organizational learning (e.g. Preskill, 1994; Preskill and Torres, 1998), as well as our own sense that future work may clarify the distinct mechanisms that underlie organizational learning, we have not emphasized organizational learning processes in Table 1. More generally, subsequent inquiry is likely to result in revisions to the framework, as some mechanisms are found to be critical in the pathways that lead to evaluation influence and others are not. In short, future conceptual and empirical work may lead to modifications of Table 1. Additionally, the general framework could be tailored to specific contexts.<sup>10</sup>

A second caveat is that, because of space limits, we have not focused adequately on the various complexities that impinge on evaluation influence processes. In part, these complexities are represented by the 'Decision/policy setting' box in Figure 2. For example, evaluation takes place within a marketplace of competing information and values. Other forces, such as program clients, advocacy organizations and partisan politicians often engage in influence processes that pull in a different direction than evaluation. In part, these complexities are also represented by the 'Contingencies' box in Figure 2. All change processes are contingent, in that they will operate in some circumstances and not others. For example, elaboration usually will not occur unless the person is both motivated to elaborate (e.g. because the issue is relevant to the person) and able to do so (e.g. has the time) (Petty and Cacioppo, 1986). By acknowledging such contingencies, evaluators may be more appropriately modest in their aspirations for evaluation as a source of influence that may contribute to social betterment.

Four other caveats warrant brief mention. (1) For the sake of simplicity, we have used language suggesting that evaluation influence is triggered by *an* evaluation. As growing attention to meta-analysis and other forms of synthesis suggests, evaluation's influence may often result instead from an *accumulation* of evaluation evidence. (2) Some who write about evaluation appear to have a predilection to rush to exhortations about how to ensure that evaluation makes



a difference (see Alkin [2003] on prescriptive evaluation theory). In contrast, we explicitly do not want to see the present framework translated into general prescriptions for evaluators, absent a reasonable research base to support such prescriptions. (3) Adapting an important point made by Weiss (1998), the focus should not be on *increasing* influence, but on *improving* the influence of evaluation. Improved influence, according to Weiss, would sometimes mean ignoring a seriously flawed evaluation, or relying on a meta-analysis rather than the most recent or local evaluation. (4) We have elsewhere suggested that Russian stacking dolls are a better metaphor than the ‘black box’ for studying mechanisms. That is, talking about a mechanism opens up the possibility of shifting to a more microscopic mechanism that underlies it, or to a more macro mechanism that may subsume it (e.g. one could talk about the more micro neurobiological processes that underlie elaboration, or to a more general process such as ‘thinking’ or ‘learning’ that could subsume it). Although such matters are the basis for interesting academic arguments and sometimes have practical implications, we hope that they will not excessively dominate future discussions of the mechanisms that underlie evaluation influence.

### ***Benefits of the Proposed Framework***

The potential benefits of the proposed framework include:

1. serving as a stimulus and guide to research on evaluation;
2. clarifying the nature of evaluation influence with respect to mediating processes, outcomes and level of analysis;
3. providing a potentially better guide to practice; and
4. furthering discussion about the responsibilities of the evaluator.

First, regarding research on evaluation, systematic inquiry about the consequences of evaluation is essential in order to establish an empirical basis for the effective practice of evaluation (Shadish et al., 1991). Without an empirical basis that provides evidence about what kinds of evaluation worked in which circumstances to produce which specific outcomes, the field of evaluation will forever be susceptible to fads, and debates about evaluation approaches will rarely rise above ideology (Mark, 2001, 2003). Simply having a new framework may help stimulate more research. The set of processes summarized in Table 1 should be especially conducive to stimulating more and better research because they should facilitate evaluators’ awareness of, interest in, and learning from other research literatures that address change processes at individual, interpersonal and organizational levels. Evaluators have largely failed to incorporate into their taxonomies of use the myriad theories of change processes and the associated research bases in the social and behavioral sciences (though important exceptions exist, including Weiss’s [1988] attention to the literature on decision processes within organizations, Preskill’s [1994] and others’ growing consideration of work on organizational learning, and Nutley et al.’s [2002] recent synthesis of lessons from the diffusion of innovation literature). Established research literatures can be a powerful source of methods, measures and hypotheses. Additional benefits (e.g. cross-pollination of ideas between evaluation and other

### *Evaluation 10(1)*

fields) may also accrue if evaluation starts to connect better with these other research areas.

Second, the framework offered here can provide clarity that is lacking in extant models of use. Existing notions of use are too imprecise, too oriented toward end-states, and too ambiguous regarding level of analysis to fruitfully guide the next generation of research and practice. For example, although it has been quite valuable, the popular notion of enlightenment (or conceptual use) is more a metaphor than a well-specified and well-operationalized concept. A shift to research-based change processes such as those in Table 1 can go a long way toward providing greater clarity. The framework also overcomes a shortcoming of prevalent models of use, which inappropriately direct attention away from the immediate effects and processes that may (or, in some circumstances, may not) lead to eventual use. Standard taxonomies of types of use instead tend to focus on end-states such as instrumental use and conceptual use. But the steps between evaluation and end-states such as policy change may be long and arduous. As noted earlier, without attention to the individual steps in the pathways, understanding of evaluation influence will be unnecessarily limited. Moreover, it is as important to identify the conditions under which certain pathways of influence do *not* lead to an end-state such as policy change, as it is to know whether that end-state has occurred in a specific case. The framework offered here is also clear as to the level of analysis, unlike existing taxonomies of evaluation use. For instance, the pathway leading from evaluation to changes in an individual teacher's instructional practices is likely to be very different from the pathway leading to a state legislature's funding of a new educational program (Weiss, 1998). The framework of evaluation influence, represented in Table 1 and Figure 2, clearly distinguishes between the two types of behavioral influence at different levels of analysis, unlike the current concept of instrumental use.

Third, the greater clarity of the framework makes it a better guide to practice. For example, by combining change processes into planned influence pathways, evaluation practitioners can develop a more comprehensive plan for maximizing the influence of a specific evaluation. By reviewing the various processes in Table 1, practicing evaluators may identify change mechanisms, such as salience or agenda setting, which they may not otherwise have targeted. Indeed, even simply reviewing the four general types of processes (general influence, cognitive and affective, motivational and behavioral) may suggest additional activities for a dissemination and influence plan. For instance, are there motivational levers that should be pulled, in an effort to increase the possible influence of evaluation findings on practitioners' behavior? The framework's distinction among the levels of analysis may also contribute to practice: evaluators can readily consider whether they should be trying to facilitate individual, interpersonal or collective processes, or some combination. In addition, the model in Figure 2 can also contribute to evaluation practice. For example, Figure 2 posits that, if evaluation knowledge production attributes have any effect, it is because they help trigger one or more evaluation mechanisms and processes. Thus, the present model suggests that, rather than focus solely on attributes such as responsiveness and credibility, evaluation practitioners may do better if they attend to how these

attributes can facilitate underlying change mechanisms. In these and other ways (including, eventually, the possibility of better research on influence), the framework presented in this article can serve as a tool to help achieve more evaluation influence in practice.

Fourth, the framework offered here could perhaps contribute to more appropriate expectations about the evaluator's responsibilities for the effects of an evaluation. Evaluators have greater control over some forms of influence than others. With the present framework of evaluation influence, the evaluation community may be better able to discuss evaluators' responsibility for evaluation influence. For instance, it may be more appropriate for evaluators to be held responsible for selected short-term processes within complex influence pathways, rather than to hold them responsible for a traditional end-state use as some normative treatments of use appear to do. Such a shift could actually enhance accountability, by setting reasonable and measurable expectations for evaluators prior to the start of an evaluation.

### **Summation**

In this article we have presented a framework for classifying the mechanisms that may underlie evaluation influence. The framework includes three levels of analysis (individual, interpersonal and collective) and, within each level, four kinds of processes (general influence, cognitive and affective, motivational and behavioral). These different levels and types of processes have generally been left confounded or undifferentiated in prior work on use. The specific processes in the framework can be combined, within and across levels (e.g. from individual to interpersonal to collective), to represent the complex pathways through which evaluation can contribute to change. We have also sketched out a broader model of evaluation influence, in which the underlying processes are located in the context of evaluation inputs, evaluation activities, the knowledge attributes of evaluation findings, and various sources of contingencies. Taken together, these pieces can help stimulate a new generation of thinking about evaluation's effects, thereby better guiding research, theory and practice.

### **Notes**

This is adapted from a paper presented at the 5th biennial meeting of the European Evaluation Society, 12 October 2002 in Seville, Spain. The authors thank Elliot Stern, Peter Dahler-Larsen and three anonymous reviewers for their helpful comments on previous versions of this article. Work on this article was supported in part by a grant from NSF (REC-02311859) to Melvin M. Mark.

1. The term 'use' is often treated interchangeably with the term 'utilization' (Weiss, 1993). These two terms have been applied in nearly synonymous ways in the evaluation literature and we do not here attempt to distinguish between the two.
2. Technically speaking, the outputs of an evaluation are the reports, presentations, briefings and other 'products' of the evaluation. The production attributes that Cousins refers to, which are drawn from the literature on use, are in a sense judgments made by stakeholders about the products of the evaluation.

## *Evaluation 10(1)*

3. As an example, in the US the evaluation of Project High/Scope provided empirical justification for Head Start and other compensatory early education programs far beyond Ypsilanti, Michigan, the site of the original study (Barnett, 1992). The Project High/Scope evaluation has been an important influence on US public policy, but these effects would be omitted for consideration in a study of evaluation use if Alkin's limited scope definition of evaluation use is employed. Regardless of whether or not this example meets a definition of 'evaluation use' (or is instead a case of 'knowledge use', as Alkin would classify it), it is a distinct and important form of evaluation influence and should not be overlooked.
4. By 'influence processes' we are referring to important *mechanisms* through which change occurs. The more general term 'evaluation influence' refers to the changes that occur as a result of influence. To use a grammatical analogy, 'influence processes' are active verbs, and 'evaluation influence' is a noun.
5. In fact, mechanisms are included in the general influence category in large part precisely because their role is in stimulating other processes that lead to (or away from) social betterment.
6. We were stimulated to add the motivational category by recent work by Walter et al. (2003a, 2003b). We thank an anonymous reviewer for pointing us toward this unpublished work.
7. A descriptive norm is a belief about which behaviors others typically perform, while an injunctive norm is a belief about which behaviors others typically approve or disapprove of.
8. The framework may also help illuminate differences between evaluation theories. To take but one example, we expect that the typical participatory and collaborative evaluator will be concerned primarily with the interpersonal column of Table 1, and may be more concerned with motivational processes than many other evaluators.
9. We assume that one of the general influence processes, perhaps elaboration, was responsible for stimulating this increased salience, but Melkers and Willoughby do not provide data that allow us to discern this initial step in the influence pathway.
10. While revising this article, we discovered an additional reason for the caveat cautioning against seeing the current framework as the final word in the mechanisms that underlie evaluation influences. We encountered work by Walter et al. (2003a, 2003b) offering their views of the 'mechanisms through which [interventions in the literature] aim to enhance research impact' (2003b: 12). The eight mechanisms that Walter and her colleagues identify are: dissemination, educational interventions, social influence, collaborations between researchers and users, incentives, reinforcement of behavior, facilitation and multifaceted interventions that use two or more of the preceding. We believe that several of these map fairly closely onto parts of our framework in Table 1. We see others as kinds of interventions, or aspects of the decision/policy setting, rather than as underlying mechanisms. Nevertheless, we take some pleasure in the prospect that the field may soon have competing models of the mechanisms that may underlie evaluation influence, rather than no such models.

## **References**

- Alkin, M. C. (2003) 'Introduction (to Section on Evaluation Use)', in T. Kelligan and D. L. Stufflebeam (eds) *International Handbook of Educational Evaluation*, pp. 189–96. Dordrecht: Kluwer Academic Press.
- Alkin, M. C., R. Daillak and P. White (1979) *Using Evaluations: Does Evaluation Make a Difference?* Beverly Hills, CA: Sage.

*Mark and Henry: The Mechanisms and Outcomes of Evaluation Influence*

- Barnett, W. S. (1992) 'Benefits of Compensatory Preschool Education', *Journal of Human Resources* 27(2): 279–312.
- Caplan, N. (1977) 'A Minimal Set of Conditions Necessary for the Utilization of Social Science Knowledge in the Policy Formulation at the National Level', in C. H. Weiss (ed.) *Using Social Research in Public Policy Making*. Lexington, MA: Lexington Books.
- Caracelli, V. J. (2000) 'Evaluation Use at the Threshold of the Twenty-first Century', in V. Caracelli and H. Preskill (eds) *The Expanding Scope of Evaluation Use. New Directions for Evaluation* 88. San Francisco, CA: Jossey-Bass.
- Cousins, J. B. (1996) 'Consequences of Researcher Involvement in Participatory Evaluation', *Studies in Educational Evaluation* 1(1): 3–27.
- Cousins, J. B. (2003) 'Utilization Effects of Participatory Evaluation', in T. Kelligan and D. L. Stufflebeam (eds) *International Handbook of Educational Evaluation*, pp. 245–66. Dordrecht: Kluwer Academic Press.
- Cousins, J. B. and K. A. Leithwood (1986) 'Current Empirical Research on Evaluation Utilization', *Review of Educational Research* 56(3): 331–64.
- Henry, G. T. (2000) 'Why Not Use?', in V. Caracelli and H. Preskill (eds) *The Expanding Scope of Evaluation Use. New Directions for Evaluation* 88, pp. 85–98. San Francisco, CA: Jossey-Bass.
- Henry, G. T. and M. M. Mark (2003) 'Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions', *American Journal of Evaluation* 24(3): 293–314.
- Hofstetter, C. H. and M. C. Alkin (2003) 'Evaluation Use Revisited', in T. Kelligan and D. L. Stufflebeam (eds) *International Handbook of Educational Evaluation*. Dordrecht: Kluwer Academic Press.
- King, J. A. (2002) 'Building the Evaluation Capacity of a School District', in D. W. Compton, M. Baizerman and S. H. Stockdill (eds) *The Art, Craft, and Science of Evaluation Capacity Building. New Directions for Evaluation* 93, pp. 63–80. San Francisco, CA: Jossey-Bass.
- Kirkhart, K. (2000) 'Reconceptualizing Evaluation Use: An Integrated Theory of Influence', in V. Caracelli and H. Preskill (eds) *The Expanding Scope of Evaluation Use. New Directions for Evaluation* 88, pp. 5–24. San Francisco, CA: Jossey-Bass.
- Knorr, K. D. (1977) 'Policymakers' Use of Social Science Knowledge: Symbolic or Instrumental?', in C. H. Weiss (ed.) *Using Social Research in Public Policy Making*. Lexington, MA: Lexington Books.
- Mark, M. M. (2001) 'Evaluation's Future: Furor, Futile, or Fertile?', *American Journal of Evaluation* 22(3): 457–79.
- Mark, M. M. (2003) 'Toward a Comprehensive View of the Theory and Practice of Program and Policy Evaluation', in S. I. Donaldson and M. Scriven (eds) *Evaluating Social Programs and Problems: Visions for the New Millennium*, pp. 183–204. Hillsdale, NJ: Erlbaum.
- Mark, M. M., G. T. Henry and G. Julnes (2000) *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs*. San Francisco, CA: Jossey-Bass.
- Melkers, J. E. and K. G. Willoughby (1998) 'The State of the States: Performance-Based Budgeting Requirements in 47 Out of 50', *Public Administration Review* 58(1): 66–73.
- Melkers, J. E. and K. G. Willoughby (2001) 'Budgeters Views of State Performance-budgeting Systems: Distinctions across Branches', *Public Administration Review* 61(1): 54–64.
- Nutley, S., H. Davies and I. Walter (2002) 'Conceptual Synthesis 1: Learning from the Diffusion of Innovations'. Unpublished manuscript, University of St Andrews, Research Unit for Research Utilization.

*Evaluation 10(1)*

- Patton, M. Q. (1997) *Utilization-focused Evaluation: The New Century Text*. Thousand Oaks, CA: Sage.
- Patton, M. Q., P. S. Grimes, K. M. Guthrie, N. J. Brennan, B. D. French and D. A. Blyth (1977) 'In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research', in C. H. Weiss (ed.) *Using Social Research in Public Policy Making*. Lexington, MA: Lexington Books.
- Pawson, R. and N. Tilley (1997) *Realistic Evaluation*. London: Sage.
- Petty, R. E. and J. T. Cacioppo (1986) *Communication and Persuasion: Central and Peripheral Routes to Attitude*. New York: Springer-Verlag.
- Preskill, H. (1994) 'Evaluation's Role in Enhancing Organizational Learning', *Evaluation and Program Planning* 17(3): 291–7.
- Preskill, H. and V. J. Caracelli (1997) 'Current and Developing Conceptions of Use: Evaluation Use Topical Interest Group Survey Results', *Evaluation Practice* 18(3): 209–25.
- Preskill, H. and R. Torres (1998) *Evaluative Inquiry for Organizational Learning*. Thousand Oaks, CA: Sage.
- Rog, D. J. (1985) *A Methodological Analysis of Evaluability Assessment*, Doctoral dissertation. Nashville, TN: Vanderbilt University.
- Shadish, W. R., T. D. Cook and L. C. Leviton (1991) *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: Sage.
- Shula, L. M. and J. B. Cousins (1997) 'Evaluation Use: Theory, Research, and Practice since 1986', *Evaluation Practice* 18(3): 195–208.
- Walter, I., S. Nutley and H. Davies (2003a) 'Developing a Taxonomy of Interventions Used to Increase the Impact of Research'. Unpublished manuscript, University of St Andrews, Research Unit for Research Utilization.
- Walter, I., S. Nutley and H. Davies (2003b) 'Research Impact: A Cross Sector Literature Review'. Unpublished manuscript, University of St Andrews, Research Unit for Research Utilization.
- Weiss, C. H., ed. (1977) *Using Social Research in Public Policy Making*. Lexington, MA: Lexington Books.
- Weiss, C. H. (1979) 'The Many Meanings of Research Utilization', *Public Administration Review* 39(Sept/Oct): 426–31.
- Weiss, C. H. (1988) 'Evaluation for Decisions: Is Anybody There? Does Anybody Care?', *Evaluation Practice* 9(1): 5–20.
- Weiss, C. H. (1998) 'Improving the Use of Evaluations: Whose Job is it Anyway?', *Advances in Educational Productivity* 7: 263–76.
- Weiss, C. H. and M. J. Bucuvalas (1977) 'The Challenge of Social Research to Decision Making', in C. H. Weiss (ed.) *Using Social Research in Public Policy Making*. Lexington, MA: Lexington Books.
- Williams, K., B. de Laat and E. Stern (2002) *The Use of Evaluation in the Commission Services: Final Report*. Paris: Technopolis France.
- Willoughby, K. G. and J. E. Melkers (2001) 'Assessing the Impact of Performance Budgeting: A Survey of American States', *Government Finance Review* 17(2): 25–30.

MELVIN M. MARK is Professor of Psychology at the Pennsylvania State University and Editor of the American Journal of Evaluation. Please address correspondence to: Department of Psychology, Penn State, 407 Moore, University Park, PA 16802, USA. [email: m5m@psu.edu]

*Mark and Henry: The Mechanisms and Outcomes of Evaluation Influence*

GARY T. HENRY is a Professor in the Andrew Young School of Policy Studies and Department of Political Science at Georgia State University. Please address correspondence to: MSC 2A 1230, Georgia State University, 33 Gilmer Street Unit 2, Atlanta, GA 30303–3082, USA. [email: dpogth@langate.gsu.edu]