Articles should deal with topics applicable to the broad field of program evaluation. Implications for practicing evaluators should be clearly identified. Examples of contributions include, but are not limited to, reviews of new developments in evaluation and descriptions of a current evaluation effort, problem, or technique. Manuscripts should include appropriate references and normally should not exceed 10 double-spaced typewritten pages in length; longer articles will occasionally be published, but only where their importance to EP readers is judged to be quite high.

Evaluation Use: Theory, Research, and Practice Since 1986

LYN M. SHULHA and J. BRADLEY COUSINS

Lyn M. Shulha

ABSTRACT

As part of a larger effort by members of the American Evaluation Association (AEA) Topical Interest Group on Evaluation Use (TIG-EU), we undertook an extensive review and synthesis of literature in evaluation use published since 1986. We observe several recent developments in theory, research and practice arising from this literature. These developments include: the rise of considerations of context as critical to understanding and explaining use; identification of process use as a significant consequence of evaluation activity; expansion of conceptions of use from the individual to the organization level; and diversification of the role of the evaluator to facilitator, planner and educator/trainer. In addition, understanding misutlilization has emerged as a significant focus for theory and to a limited extent, research. The article concludes with a summary of contemporary issues, particularly with regard to their implications for evaluation practice.

INTRODUCTION

The annual meeting of the American Evaluation Association (AEA) in 1996 marked not only the 10th anniversary of the Association itself (from an amalgamation of the Evaluation Net-

Lyn M. Shulha • Faculty of Education, Queen's University, Kingston Ontario K7L 3N6, Canada; **J. Bradley Cousins** • University of Ottawa, P.O. Box 450 Stn. A, Ottawa Ontario K1N 6N5, Canada.

Evaluation Practice, Vol. 18, No. 3, 1997, pp. 195–208.
ISSN: 0886-1633

work and the Evaluation Research Society), but also a decade of activity by evaluation theorists, researchers, and practitioners interested in the nature, causes, and consequences of evaluation utilization.[1] A conceptual home for these people has been the Topical Interest Group in Evaluation Use (TIG-EU). The anniversary occasion prompted TIG-EU to lead a presidential strand session titled, "The past, present, and future of evaluation use: The continuing evolution of evaluation practice and theory." The session was designed to take stock of current understandings and practices as they relate to use, and to help orient the research agenda. Coincidentally, no major review and synthesis of literature in the field had been published since 1986. In preparation for the presidential session, we carried out such a review. A portion of that paper is summarized here with a particular emphasis on implications for evaluation practice.

The literature reviewed was published in the period 1986 to the present. The temporal boundaries of this investigation are not meant in any way to diminish the significance of earlier work in the field. In fact, we direct interested readers to several now-classic works in the field (e.g., Alkin, Daillak, & White, 1979; Alkin, Koescoff, Fitz-Gibbon, & Seligman, 1974; Patton, et al., 1977; and Weiss, 1973, 1977, 1979). This work, we believe, has significantly shaped evaluation utilization as a field of inquiry. These early efforts and those of other researchers and theoreticians soon led to an accumulation of conceptual and empirical literature so extensive that, by the early 1980's, several comprehensive reviews and syntheses of what had been learned appeared. Seven such reviews[2] provided the backdrop for our own. We begin by rapidly describing this baseline and then proceed to thematically work our way through several recent developments in the literature. Our goal is to extrapolate from such themes implications for practicing evaluators.

STARTING POINTS FOR A NEW DECADE OF RESEARCH

By 1986 there was general agreement that use, rather than being a unitary concept, was a multi- dimensional phenomenon best described by the interaction of several dimensions, namely, the instrumental (decision support and problem solving function), conceptual (educative function), and symbolic (political function) dimensions (see Leviton & Hughes, 1981 for a discussion of these dimensions). In an effort to enhance evaluation use along these dimensions, survey articles produced an extended "shopping list" of likely predictors. These predictors tended to cluster into categories such as: (a) relevance, (b) credibility, (c) user involvement, (d) communication effectiveness; (e) potential for information processing; (f) clients' need for information; (g) anticipated degree of program change; (h) perceived value of evaluation as a management tool; (i) quality of evaluation implementation; and (j) contextual characteristics of the decision or policy setting. Cousins and Leithwood (1986) endeavored to move beyond merely describing the factors influencing utilization by developing a meta-analytic method to assess the relative weight of factors in their ability to predict use. Their findings indicated that the quality, sophistication, and intensity of evaluation methods were among the most potent in influencing the use of these findings.

Utilization research to this point had managed to highlight the complexity of use, and by implication, the complexity of the program and policy contexts in which evaluation occurred. It was acknowledged that the design of useful evaluations required much more than simply matching clients' questions with particular methods, approaches or models. More and more, evaluators were taking great pains to identify the type of use sought by program stakeholders

and to custom-design evaluations that would best promote such uses. The emergence of a conception of evaluation as a continuous information dialogue began to lead to more frequent occurrences of evaluators and program stakeholders sharing responsibility for generating, transmitting, and consuming evaluation information. Such findings contributed to a resurgence of enthusiasm for the potential of evaluation and a new decade of conceptual and empirical research. It is on this new era that we now focus attention.

THEMES OF DEVELOPMENT IN EVALUATION UTILIZATION

The Rise of Context in Utilization Theory

The preliminary meta-analytic work of Cousins and Leithwood (1986), and later, direct observations by Greene (1988), challenged the research community to attend to the task of theory building. They did so by arguing that if one's goal is to promote utilization, it is not enough simply to describe different types of use and to catalogue the contributing factors. Since that time, there have been only tentative steps in this direction. For example, Johnson (1993) invoked a "meta-modeling" procedure to integrate and synthesize several different utilization conceptual frameworks. While this study helped to clarify the nature and inter- relational properties of variables explaining use, and indeed of the utilization construct itself, it fell short of specifying the relative weighting of influential factors. In a separate study, Levin (1987) applied the framework developed by Cousins and Leithwood (1986) to empirically test the relative weight of factors influencing different types of use. His findings contradicted those of Cousins and Leithwood by underscoring the pivotal nature of contextual factors in explaining patterns of use.

Arguably the most influential conceptual contributions to theory development emerged from the well- known "Weiss-Patton debate" (Patton, 1988a; Weiss, 1988a, 1988b). This exchange, which was instigated by Weiss' provocative comments at the 1987 AEA annual meeting in Boston, continued publicly in *Evaluation Practice* for the benefit of all evaluators. At the heart of the matter was the question of whether evaluators could, or should be held accountable for use. At the risk of oversimplifying, Weiss argued that evaluation information competes for credibility in decision-making contexts best characterized as non- rational and pluralistic. In these contexts, she claimed, the most an evaluator can and should strive for is the generation of evaluation findings that are accurate and adequate for the question. Patton, on the other hand, advocated an active role for evaluators in promoting and cultivating use. Evaluators, he maintained, had responsibilities to (a) help decision makers identify their evaluation needs, (b) determine, in consultation, what information would best address those needs, and (c) generate findings that would yield the type of information needed by the intended user. While a clear resolution of this issue was not readily forthcoming, Smith and Chircop (1989) acknowledged the significance of the debate to the profession and pointed out that each position was argued out of different program contexts and with the interests of different stakeholders in mind. A theory of use, they suggested, will likely differentiate between contexts where stakeholders debate the goals of the program (political context) and those in which the goals are agreed upon but the means to reach these goals are the focus for inquiry (rational decision making context).

Concern for the conditions underlying evaluation marked a growing effort by researchers and theorists to explicate the role of context in utilization. Collectively, what ensued was

a multifaceted investigation with epistemological, political, and structural components. First, some prominent evaluation theorists (e.g., Guba & Lincoln, 1989) challenged traditional assumptions about the nature of knowledge. In advocating a constructivist approach to evaluation, as they did, the creation of knowledge is conceived to be integral to the context in which it is produced. Thus, because evaluations create knowledge, the very act of evaluating is use (Finne, Levine & Nilssen, 1995).

A second perspective argued that political activity is inextricably linked to context—the close connection of people, environment, and circumstances—which, in turn, influences use. From their own experiences, Greene (1990) and King (1988) added strength to Weiss' (1988a, 1988b) contentions that decision makers do not act alone and face an onslaught of decision-relevant information from competing interest groups. On this point Thompson (1994) reminded us of Herbert Simon's work in organization theory, particularly as it relates to pressure on administrators to engage in 'satisificing' behavior (i.e., deciding on a course of action that is good enough while still leaving open options to change course should that become necessary). Shulha and Wilson (1995), incidentally, illustrated such behaviors in the context of collaborative evaluation. Also using political frames of reference, other researchers have described how the loss or acquisition of resources during an evaluation (usually due to unanticipated and politically charged events) changed in significant ways the effects of the evaluation (Mowbray, 1992; Patton, 1988b; Tomlinson, Bland & Moon, 1993; Weiss, in Alkin, 1990).

A growing familiarity with the pivotal nature of program contexts was also evident in a third stream of inquiry that focused on the relationship between use and organizational structures and processes (Jenlink, 1994; Mathison, 1994; Owen & Lambert, 1995; Preskill, 1991, 1994a; Preskill & Torres, 1996; Torres, Preskill & Piotnek, 1996). The complexity of bureaucratic levels within the organization, the lines of communication within and across these levels, and the dominant mechanisms for framing the meaning of evaluation information all contributed to the potential utility of evaluation findings.

Several researchers and theorists who were intrigued by the potency of contextual factors focused more directly on the interaction between the evaluator and the program context. Patton (1988a, 1994, 1997) has been a key contributor to this line of inquiry. In arguing that evaluations must serve the intended use of intended users, Patton positions the evaluator in the thick of program context. For Patton, "situation recognition and responsiveness, anticipation, and being able to analyze people" (1988a, p. 19) are the basis for dialogue, deliberation and utilization-focused designs. This issue of connectedness of evaluators to program contexts has captured the attention of evaluation researchers and theorists interested in social and psychological perspectives on human learning. Several authors, for example, acknowledged the power of socially constructed meanings of evaluation information (Cousins & Earl, 1992; Finne et al., 1995; Marsh & Glassick, 1988; Owen, Lambert, & Stringer, 1994). Evaluation provides a forum for the social processing of information and the development of shared interpretations of it. Not all theorists have concurred with the collective learning perspective, however (e.g., Huberman, 1995). Arguing from a psychological perspective, Huberman and Cox (1990) encourage evaluators to find ways to connect cognitively with their clients so as to frame findings in ways that are inherently meaningful to these users. In addition to the user-learning argument, be it sociological or psychological, tighter connections between the evaluator and the program community have also been defended on the basis of the insights they yield about the needs of decision makers and the program community, and the implications of such insights for enhancing evaluation's responsiveness (Alkin, 1991; Brown-McGowan & Eichelberger, 1993; Huberman & Cox, 1990).

EXPANDING CONCEPTIONS OF USE

Some of the more innovative thinking about use arises from efforts to understand the impact of the evaluation process quite apart from the nature of findings. Greene (1988) developed a grounded conceptual framework for use that is very much tied to process. The cognitive, affective, and political dimensions of this framework proved to be robust in anchoring subsequent research on process use. For example, several studies have closely examined the participation of intended users in the evaluation process, the impact of such participation on personal learning, and the ultimate effect this learning has on program practices (Ayers, 1987; Cousins, 1995, 1996; Patton, 1994; Preskill, 1994;). Similarly, others have observed how participation in evaluation gives stakeholders confidence in their ability to use research procedures, confidence in the quality of the information that is generated by these procedures, and a sense of ownership in the evaluation results and their application (Ayers, 1987; Cousins, 1995; Greene, 1988). Finally, engagement, self determination, and ownership appear as essential elements in evaluation contexts where stakeholders either anticipate, or are immersed in deliberate political action concerning their programs (Fetterman, 1994; Fetterman, Kaftarian & Wandersman, 1996).

The most recent addition to process use literature comes from Patton (1997). His conceptual perspective is based on the theory that process use can lead to four different, but interrelated consequences. First, it can enhance communication within an organization by sending a clear message to the community about what elements of program are important to investigate. Second, data collection can take the form of interventions with the purpose of improving the program while examining it. Third, the evaluation processes can be used to engage program participants in their work in even more crucial ways, giving them opportunities to be reflective and discriminating about the type of information most appropriate to their efforts. Finally, the process can lead not only to individual, but to organizational development.

Investigations of process use throughout the decade have looked beyond effects on individual participants. As the acknowledged inter-relationships between evaluation and program context became more acute, so did interest in the potential of evaluation to foster growth in the program community as a whole. The notion of evaluation for organizational learning has attracted considerable attention. For example, Cousins and Earl (1992) argued for a pragmatic form of participatory evaluation using principles of organizational learning as a theoretical rationale.[3] Subsequently, several theorists made strong cases for understanding evaluation impact in an organizational context (Jenlink, 1994; Owen & Lambert, 1995; Preskill & Torres, 1996). Jenlink (1994) was especially interested in how the evaluator might unearth operative framing mechanisms so that organization members might see and understand their collective behaviors more clearly.

While empirical evidence supporting the power of evaluation to foster organizational learning is mixed (Cousins & Earl, 1995; Forss, Cracknell & Samset, 1994), a significant body of work has shown signs of relationship between evaluation activities and the development of organizational capacity (Cousins, 1995, 1996; Preskill, 1994; Timar, 1994; Torres et al., 1996). Interest in this expanded notion of evaluation use appears to be gaining momentum. As Timar (1994, p.65) comments, "The challenge to policy makers is to design programs and evaluation in a way that encourages organizational coherence and integration rather than fragmentation."

Emergent Evaluation Practices

There is little doubt that emphases on contextual understandings and process use have ties to emergent evaluation practices. Most noticeably we observe the proliferation of collaborative modes of evaluation. Unlike strictly consultative approaches, collaborative models tend to aspire to more equitable power relationships between evaluators and program practitioners leading to jointly negotiated decision making and meaning making.

What has been learned about the complexities and effects of collaboration? Several studies have reported that evaluators working in partnership with stakeholders are better able to stimulate knowledge production and promote positive reactions to evaluation findings (Ayers, 1987; Cousins, 1995; Cousins & Earl, 1995; Greene, 1988). Yet countervailing evidence raises concerns about the clash between evaluator and program practitioner cultures and formidable obstacles to genuine program stakeholder participation. Teachers, for example, protest the chronic lack of time that can be freed up to invest into admittedly worthwhile and relevant professional activities (Alvik, 1995; Cousins & Earl, 1995; Nevo, 1993).

Collaboration has been promoted as a vehicle for capacity building, organizational learning (Cousins & Earl, 1992, 1995) and stakeholder empowerment (Fetterman, 1994). Finne et al. (1995) detail and critique an interesting approach they call "trailing research." This collaborative model espouses (a) maintenance of a continual focus on learning, (b) clarification of roles and expectations, (c) creation of arenas for dialogue, (d) joint focus on all issues being investigated, (e) attention to the validation of findings, and (f) joint interpretation of results. While such models have enormous potential for impact, they pose formidable challenges for evaluators to implement.

Concomitant with the rise of collaborative models and program theoretic and developmental approaches to evaluation (e.g., Chen, 1994; Patton, 1994) is an expanding conception of the evaluator's role. A variety of new roles and responsibilities are part and parcel of a heightened demand for what Shadish and Epstein (1987) dubbed the "stakeholder-service orientation." The conception of evaluator as planned change consultant or agent is an increasingly common one, especially among evaluators interested in fostering process use or working toward the development of organizational learning capacity (Cousins & Earl, 1995; Goering & Wasylenki, 1993; Mathison, 1994; Owen & Lambert, 1995; Preskill, 1994; Torres et al., 1996). A related variant arises from the integration of program planner and program evaluator roles (Patton, 1994, 1997).

In each of these roles, evaluators forge partnerships within the community of program practice while at the same time taking measures to guard against idealism about the program, co-optation, and systematic bias. In many instances evaluators must relinquish control over decisions requiring substantive program knowledge and knowledge about context, while maintaining authority over the technical quality of the evaluation (Cousins & Earl, 1992; Greene, 1988). The extent to which evaluators are able to guard against undesirable influences and biases is a matter of serious concern (e.g., Scriven, 1996) and has prompted many theorists to recommend that collaborative models be limited to formative, improvement-oriented contexts (e.g., Ayers, 1987; Cousins & Earl, 1992, 1995). This is not true, however, of all advocates of collabortive modes of inquiry (e.g., Fetterman, 1994; Fetterman et al., 1996).

With these new roles comes a new mix of skills and expertise. Beyond communication skills, evaluators require "people skills" (Ayers, 1987) that allow them to build bridges and connections, especially with those skeptical of their presence. Goering and Wasylenki (1993) argued that substantive program expertise is also an asset because it facilitates the evaluator's

reconciliation of a host of program and organizational context variables that impact the use of evaluative information. These observations were directly challenged by Champney (1993), however, who pointed out that the very substantive knowledge that allows the evaluator to facilitate data collection and analysis also significantly shapes the evaluation outcomes.

Other skills finding their way into the evaluator's repertoire are those associated with the role of educator/trainer. Honed teaching skills and knowledge of adult education principles are likely to benefit those using participatory designs since engaging program practitioners in technical evaluation activities is key to their implementation (Cousins, 1996; Cousins & Earl, 1992). In some collaborative contexts, primary users of evaluation are expected to take charge of technical research responsibilities, leaving the evaluator in the background to act primarily as guide or facilitator to the process (Fetterman, 1994; Fetterman et al., 1996). Murphy (1996), who analyzed the consequences of having participants govern the many elements of evaluation, reported (among other things) that the quality of participants' training was essential to both immediate evaluation use and their willingness to re-engage in evaluation activities.

Among the more fundamental role changes for evaluators interested in collaborative models is their affinity for establishing partnerships in the first place. As Shulha (1994, p. 6) lamented, "Wishing an evaluation to be participatory, does not make it so!" And King (1995) pointed out that while ideal participants are both willing and able, one of these attributes may often be missing. Several scholars have remarked that the likelihood of inculcating a penchant for systematic inquiry into organizations would occur only in concert with concomitant changes in organizational culture (Cousins & Earl, 1995; Keiny & Dryfus, 1993; Murphy, 1996; Nevo, 1993). Leadership in cultural change often falls to program administrators and, as clearly illustrated by Keiny and Dryfus (1993), it is often the willingness of management to support program participants (as well as to accept, appreciate, and use evaluative information) that give credibility to participation.

Attention to Misuse

As we openly acknowledge, the swing toward collaborative models raises serious questions about the evaluator's ability to maintain a sufficiently bias-free stance due to pressures emanating mainly from within the program community. While being undeniably important in any evaluation context, the question of understanding the complexities of, and the potential for, patterns of misutilization is especially pivotal in situations where evaluators work closely with program stakeholders. Perhaps for this reason, researchers and theorists interested in evaluation use have taken on the challenge of developing our understanding of the concept of misutilization more than any other related field of inquiry (e.g., knowledge utilization, research utilization).

Framing Misutilization. Among the first notable attempts to conceptually differentiate questionable practices were Alkin and Coyle (1988), who distinguished among *justified non-use* of evaluation data, mischievous use (*misuse*) of evaluation, *misevaluation* (i.e., the use of data of dubious quality), and *abuse* or the suppression of quality information for political or otherwise covert reasons. Patton (1988b) also helped develop thinking in this domain. In his framework, the first contrast is between utilization and non-utilization of evaluation findings; a second dimension places" "non-misutilization" (justified use) in opposition to "misutilization." Patton argued that one reason to distinguish these two dimensions is that the methods used to study misuse would likely prove to be different than those currently popular

for researching utilization. A variable for which one would need to account in misutilization research, for example, would be "intentionality." Several examples of misutilization have surfaced in the literature. These include: (a) the commissioning of evaluation for purely symbolic purposes (Dial, 1994; Duffy, 1994; Stahler, 1995); (b) the conscious subvertion of the evaluation process by program practitioners (Vroom, Colombo & Nahan, 1994); and (c) the purposeful non-use of high quality information (Cousins, 1995; Palumbo, 1994).[4]

To help evaluation practitioners reflect on their own experiences, Stevens and Dial (1994) posed the question, "What constitutes misuse?" followed by a variety of arguments and scenarios of misuse. These scenarios included practices such as changing evaluation conclusions, selectively reporting results, ascribing findings to a study that differ from actual results, oversimplifying results, failing to qualify results, and accentuating negative findings. But serious systematic inquiry in this domain is problematic. While misutlization is relatively easy to define in the abstract, studying it empirically is quite another matter. In addition, the goal of developing and applying standards of behavior is hampered, as we shall see, due to ambiguity about the concept.

The Ambiguity of Misuse. The complexity of misutilization was one of the issues addressed by Alkin and a group of notable evaluation scholars during a retreat held at the UCLA Malibu Conference Facility. Their attention to this and other topics was captured in *Debates on Evaluation* (Alkin, 1990).

Weiss was one of the first scholars to caution the field against becoming too comfortable with definitive frameworks that account for misuse. As she pointed out, "... when you start taking into account the larger social good, it becomes very hard to make judgments about the immediate use and misuse of evaluation. People do rather wondrous things because there is a larger good involved" (Weiss, in Alkin, 1990, p. 254). It might be argued, then, that the more complete the understanding of the context for evaluation, the more ambiguous the case for apparent misuse becomes.

McTaggart (1991) found himself altering an intended evaluation report due to a last-minute request by the program's teacher not to use data she had initially volunteered. It became evident to the evaluator that the teacher revoked her statements because of duress placed on her by the school principal. In creating the final report, McTaggart weighed reporting what he knew to be true about the program against the likelihood that a frank report would carry with it relatively severe consequences for the teacher. It is apparent in this study, that as the evaluation progressed, the way knowledge about programs was being generated within the school changed. The principal, unaccustomed to this type of systematic inquiry and the depth of information it could produce, responded in perhaps the only way he knew how—aggressively. McTaggart's purposeful non-use of evaluation information of good quality was precipitated by two considerations: (a) the unanticipated impact of new knowledge on the power relationships within the organization, and (b) the agreement struck between the evaluator and the teacher at the beginning of the evaluation that the teacher could vet all of her comments before they were folded into the report. The evaluator, sensitive to the apparently inevitable outcome of a particularly negative report, chose to protect the teacher.

This example, and others like it, attest to the ambiguities inherent in our conceptions of misuse. It may be that evaluation practitioners who feel compelled to weigh the benefits and costs of their work within the broader context of the program's utility, or the participants' well-being, are most vulnerable to behaviors that, according to contemporary definitions, would constitute evaluation misutilization (Palumbo, 1994).

Guarding Against Misuse. As an aid to help evaluators who wish to avoid circumstances that could easily lead to misuse, evaluation theorists have put forward three complementary recommendations. They advise evaluators to seek out independent checks on their evaluation processes, to engage in methodological reviews and to consult existing codes of practice (see discussion in Alkin, 1990). On the latter point, there have been two professional resources developed in North America within the last ten years for the expressed purpose of guiding evaluation practices in normative ways. In 1994, the Joint Committee on Standards for Educational Evaluation published the well-known *The Program Evaluation Standards (2nd Edition)*. The Joint Committee articulated 30 standards of practice based on agreement within four conceptual categories: utility, feasibility, propriety, and accuracy. An independent, yet complementary document, *Guiding Principles for Evaluators*, was produced by an AEA task force (1995). Written for a full range of evaluation contexts, the five guiding principles articulated were: systematic inquiry, competence, integrity and honesty, respect for people, and responsibility for general public welfare.

Preference for one of these resources over the other might depend on the personal orientation of the evaluator. Our reading of the differences between them suggests the following. For those who frame evaluation and utilization as predominantly a rational decision support enterprise, and for those who strive to reveal through their work some fundamental truths about the programs they evaluate, the *Program Evaluation Standards* provide excellent support. For those less interested in practical guidance and more concerned with finding reference points from which they can reflect on the merits of their own work, the *Guiding Principles* may be more appropriate. Even with all these strategies, standards and guidelines in place, House (1995) reminds us that there will likely always be disputes about what constitutes good evaluation.

WHAT HAVE WE LEARNED?

It is clear, given the studies cited above, that both empirical and conceptual research on the nature, causes, and consequences of utilization has become immersed in issues of context. The evidence suggests that the more evaluators become schooled in the structure, culture, and politics of their program and policy communities, the better prepared they are to be strategic about the factors most likely to affect use. It is important to acknowledge, however, that all evaluations are political and all findings must compete for attention, even in the least complex of program or policy settings. The tasks for evaluators, as described above, are: to learn whose attention to secure; where and in what ways findings will have to compete; and how best to respond through the evaluation design. Of central importance is understanding when program goals or ends are agreed among stakeholder groups. This determination will help evaluators to frame the tasks as inherently political, or as more rational undertakings.

While instrumental, conceptual, and symbolic use remain part of the lexicon of evaluation utilization, the recognition and acceptance of process use has opened up new potential for the evaluation enterprise and created new questions for investigation. Much of the research activity around process use has tried to understand the influence of particular evaluation designs and practices on individual and organizational learning. A logical next step would be to develop ways to track these learnings, focusing on the degree to which they occur, their robustness, and their relationship to more traditional instrumental outcomes. If one accepts that knowledge is socially constructed and inextricably linked to context (Guba & Lincoln, 1989),

it follows that the investigation of these relationships should not remain solely within the domain of evaluation practitioners, researchers, or theorists. The primary users of evaluation (often program managers), looking beyond immediate information needs, may be intrigued by the longer-term professional and organizational development dimensions of newer, more participatory evaluation designs. By encouraging users to actively support these designs, evaluators will enhance the likelihood for process use (Cousins, 1995; Lafleur, 1995; Shulha & Wilson, 1995) and create learning partnerships around issues of utilization in general.

The onset of collaborative modes of evaluation may signal the desire of evaluators to effectively introduce into program communities the habits of systematic inquiry and reflective activity and associated technical, analytic, and reporting skills. As many of the studies cited above reveal, this has not always been a smooth or productive transition. This body of research helps us to understand the implications for evaluator role changes when traditional approaches to design, data collection, analysis and reporting decisions are suspended. It may be easier for evaluators to vary their practice when they consciously assume other valued social roles such as facilitator, planner, consultant, trainer, or educator. These complementary roles enable evaluators to make significant contributions to the growth and learning of program personnel and communities. But the question is, at what cost?

Related to this question is the need to pay close attention to Huberman and Cox' (1990) call to follow up on evaluation interventions. While research studies have reported—usually through relatively immediate and retrospective methods—the instrumental, cognitive, affective, and political consequences of evaluation, they typically do not track these dimensions over the long term. As a result, they fail to produce a comprehensive picture of personal/professional change in participants and cultural change in organizations. Such tracking is likely to become essential if evaluators are to be taken seriously and if they are to be joined by the program community in advocating evaluation for personal/professional and organizational learning.

King (1988) reminds us that time and money for evaluation and research on evaluation are scarce. These conditions argue for a research agenda that will continue to enhance the potential for utilization of evaluation efforts. While the current focus of research is on the interdependence of the evaluation and the practice communities, attention must also be paid to the dilemmas posed by the scholars debating evaluation misutilization. Instrumental, conceptual, and process uses must not be pursued at the cost of diminished standards or principles of practice. Evaluators and program practitioners need to further open the dialogue to better understand the ambiguities of evaluation practice. We commend the efforts of some evaluators to do just that (e.g., McTaggart, 1990; Palumbo, 1994). Such stories are likely to be our best source of data in research on misuse, though the risks to the narrators cannot be taken lightly. If "intentionality" is one of the keys in unlocking our understanding of misuse, a first step in promoting more dialogue around the "nonmisutilation/misutilization dimension" (Patton, in Alkin, 1990) would be to encourage the reporting of unanticipated decision points and unanticipated consequences of planned actions in utilization research.

Evidence and discussion surrounding the unintended consequences of practice, especially as these have surfaced in collaborative modes of evaluation, have led some to suggest that the evaluator is not capable of remaining sufficiently distanced from the program community. Our belief is that this view is misguided and unsupportable. But we need to learn much more about the conditions under which collaborative modes and non-collaborative modes of evaluation are appropriate and justifiable. Clearly attending to the connection between context and evaluation purpose is pivotal.

Having acknowledged this, we underscore that many of the authors cited above are committed to models of joint participation in making meaning, opportunities for invested authority, and the potential for long term, productive learning partnerships. While the specific skill set required to move into this realm is vast and complex, the learning – often acquired 'on the fly'—for both evaluators and program practitioners/users has considerable allure. What is clear is that the evaluation practice community needs to be versatile in order to be responsive to the needs of its clientele. If utilization research could monitor the evolution of evaluation practices for their 'goodness of fit' with the challenges faced by social and educational program communities, long-term benefits for the field would most likely result.

NOTES

This paper was adapted from a lengthy version presented at the annual meeting of the American Evaluation Association, Atlanta, November, 1996.

1. We employ the terms 'use' and 'utilization' interchangeably.
2. Alkin, et al. (1985); Burry, Alkin and Ruskus (1985); Cousins and Leithwood (1986); Beyer and Trice (1982); King and Thompson (1983); Leviton and Hughes (1981); Thompson and King (1981).
3. Key organizational learning theoretical features include: knowledge is considered socially constructed; learning occurs first at the level of the individual, then at the level of the collective; collective learning may be single-loop or incremental, or double-loop involving the questioning of fundamental assumptions about practice; learning can occur at a conceptual level (i.e., change in the range of behaviors) or in response to evaluated action or behaviors (see Cousins & Earl, 1992; Jenlink, 1994; Preskill & Torres, 1996).
4. Weiss (in Alkin, 1990), however, argues that selective use of evaluation findings is indeed consistent with the political nature of organizational decision making and, therefore, not really misuse.

REFERENCES

Alkin, M. C. (1990). *Debates on evaluation*. Newbury Park, CA: Sage.

Alkin, M. C. (1991). Evaluation theory development II. In M. W. McLaughlin, & D. C. Phillips (Eds.), *Evaluation and education: At quarter century* (pp. 91–112). Chicago, IL: University of Chicago Press.

Alkin, M. C., & Coyle, K. (1988). Thoughts on evaluation utilization, misutilization, and non-utilization. *Studies in Educational Evaluation, 14*, 331–340.

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

Alkin, M. C., Jacobson, P., Burry, J., Ruskus, J., White, P., & Kent, L. (1985). *A guide for evaluation decision makers*. Beverley Hills: Sage Publications.

Alkin, M. C., Koescoff, J., Fitz-Gibbon, C., & Seligman, R. (1974). *Evaluation and decision making: The Title VII experience*. Los Angeles: Center for the Study of Evaluation.

Alvik, T. (1995). School-based evaluation: A close-up. *Studies in Educational Evaluation, 21*, 311–343.

Ayers, T. D. (1987). Stakeholders as partners in evaluation: A stakeholder-collaborative approach. *Evaluation and Program Planning, 10*, 263–271.

Beyer, J. M., & Trice, H. M. (1982). The utilization process: A conceptual framework and synthesis of empirical findings. *Administrative Science Quarterly, 27*, 591–622.

Brown-McGowan, S., & Eichelberger, T. T. (1993). The utility of the knowledge use system in a higher education evaluation setting. *Knowledge: Creation, Diffusion, Utilization, 14*(4), 401–416.

Burry, J., Alkin, M., & Ruskus, J. (1985). Organizing evaluations for use as a management tool. *Studies in Educational Evaluation, 11*, 131–157.

Champney, T. F. (1993). Comments on the use of multiple roles to maximize the utilization of outcome research. *Evaluation and Program Planning, 16*, 335–336.

Chen, H. T (1994). A panel on theory–driven evaluation and evaluation theories. *Evaluation Practice, 15*(1), 73–74. (see also panel discussion, pp. 75–93).

Cousins, J. B. (1995). Assessing program needs using participatory evaluation: A comparison of high and marginal success cases. In J. B. Cousins, & L. M. Earl (Eds.), *Participatory evaluation in education: Studies in evaluation use and organizational learning* (pp. 55–71). London: Falmer.

Cousins, J. B. (1996). Consequences of researcher involvement in participatory evaluation. *Studies in Educational Evaluation, 22*(1), 3–27.

Cousins, J. B., Donohue, J. J., & Bloom G. A. (1996). Collaborative evaluation in North America: Evaluators' self- reported opinions, practices, and consequences. *Evaluation Practice, 17*(3), 207–226.

Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis, 14*(4), 397–418.

Cousins, J. B., & Earl, L. M. (Eds.). (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. London: Falmer Press.

Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research in evaluation utilization. *Review of Educational Research, 56*(3), 331–364.

Cousins, J. B., & Leithwood, K. A. (1993). Enhancing knowledge utilization as a strategy for school improvement. *Knowledge: Creation, Diffusion, Utilization, 14*(3), 305–333.

Duffy, B. P. (1994). Use and abuse of internal evaluation. In C. J. Stevens & M. Dial (Eds.), *New directions for program evaluation: Preventing the misuse of evaluation. 64*, (pp. 25–32). San Francisco: Jossey-Bass.

Fetterman, D. M. (1994). Empowerment evaluation. *Evaluation Practice, 15*(1), 1–15.

Fetterman, D. M., Kaftarian, A. J., & Wandersman, A. (Eds.). (1996). *Empowerment evaluation: Knowledge and tools for self assessment and accountability*. Thousand Oaks, CA: Sage Publications.

Finne, H., Levine, M., & Nilssen, T. (1995). Trailing research: A model for useful program evaluation. *Evaluation, 1*(1), 11–31.

Forss, K., Cracknell, B., & Samset, K. (1994). Can evaluation help an organization to learn? *Evaluation Review, 18*(5), 574–591.

Goering, P. N., & Wasylenki, D. A. (1993). Promoting the utilization of outcome study results by assuming multiple roles within an organization. *Evaluation and Program Planning, 16*, 329–334.

Greene, J. G. (1988). Stakeholder participation and utilization in program evaluation. *Evaluation Review, 12*(2), 91–116.

Greene, J. C. (1990). Technical quality versus user responsiveness in evaluation practice. *Evaluation and Program Planning, 13*, 267–274.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

House, E. R. (1995). Principled evaluation: A critique of the AEA guiding principles. In W. R. Shadish, D. L. Newman, M. A. Schreirer & C. Wye (Eds.), *New directions for program evaluation: guiding principles for evaluators. 66*, (pp. 27–34). San Francisco: Jossey-Bass.

Huberman, M. (1995). The many modes of participatory evaluation. In J. B. Cousins, & L. M. Earl (Eds.), *Participatory evaluation in education: Studies in evaluation use and organizational learning* (pp. 103–111). London: Falmer.

Huberman, M., & Cox, P. (1990). Evaluation utilization: Building links between action and reflection. *Studies in Educational Evaluation, 16*, 157–179.

Jenlink, P. M. (1994). Using evaluation to understand the learning architecture of an organization. *Evaluation and Program Planning, 17*(3), 315–325.

Johnson, R. B. (1993, April). *Models of evaluation utilization: A meta-modeling synthesis of the literature*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta. (Educational Resources Information Center Document Reproduction Service No. ED 358 112).

Keiny, S., & Dreyfus, A. (1993). School self-evaluation as a reflective dialogue between researchers and practitioners. *Studies in Educational Evaluation, 19*, 281–295.

King, J. A. (1988). Research on evaluation and its implications for evaluation research and practice. *Studies in Educational Evaluation, 14*, 285–299.

King, J. A. (1995). Involving practitioners in evaluation studies: How viable is collaborative evaluation in schools? In J. B. Cousins, & L. M. Earl (Eds.), *Participatory evaluation in education: Studies in evaluation use and organizational learning* (pp. 86–102). London: Falmer.

King, J. A., & Thompson, B. (1983). Research on school use of program evaluation: A literature review and research agenda. *Studies in Educational Evaluation, 9*, 5–21.

Levin, B. (1987). The uses of research: A case study in research and policy. *The Canadian Journal of Program Evaluation, 2*(1), 43–55.

Leviton, L. C., & Hughes, E. F. X. (1981). Research on the utilization of evaluations: A review and synthesis. *Evaluation Review, 5*(4), 525–548.

Marsh, D. D., & Glassick, J. M. (1988). Knowledge utilization in evaluation efforts. *Knowledge: Creation, Diffusion, Utilization, 9*(3), 323–341.

Mathison, S. (1994). Rethinking the evaluator role: Partnerships between organizations and evaluators. *Evaluation and Program Planning, 17*(3), 299–304.

Mowbray, C. T., (1992). The role of evaluation in restructuring of the public mental health system. *Evaluation and Program Planning, 15*, 403–415.

Murphy, D. F. (1996). The evaluator's apprentice: Learning to do evaluation. *Evaluation, 2*(3), 321–338.

Nevo, D. (1993). The evaluation minded school: An application of perceptions from program evaluation. *Evaluation Practice, 14*(1), 39–47.

Newman, D. L., & Cai M. (1995, November). *Teacher involvement in program evaluation and evaluation use: An empirical examination of the participatory evaluation approach*. Paper presented at the joint meeting of the Canadian Evaluation Society and the American Evaluation Association, Vancouver.

Owen, J. M., Lambert, F. C., & Stringer, W. N. (1994). Acquiring knowledge of implementation and change: Essential for program evaluators? *Knowledge: Creation, Diffusion, Utilization, 15*(3), 273–284.

Owen, J. M., & Lambert, F. C. (1995). Roles for evaluation in learning organizations. *Evaluation, 1*(2), 237–250.

Palumbo, D. J., (1994). The political roots of misuse of evaluation. In C. J. Stevens & M. Dial (Eds.), *New directions for program evaluation: Preventing the misuse of evaluation. 64*, (pp. 15–24). San Francisco: Jossey-Bass.

Patton, M. Q. (1988a). The evaluator's responsibility for utilization. *Evaluation Practice, 9*(2), 5–24.

Patton, M. (1988b). Six honest serving men for evaluation. *Studies in Educational Evaluation, 14*, 301–330.

Patton, M. Q. (1994) Developmental evaluation. *Evaluation Practice, 15*(3), 311–319.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text. (3rd Ed.)*. Newbury Park, CA: Sage.

Patton, M. Q., Grimes, P. S., Guthrie, K. M., Brennan, N. J., French, B. D., & Blyth, D. A. (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*. (pp. 141–163). New York: D. C. Heath and Company.

Preskill, H. (1991). The cultural lens: Brining utilization into focus. In C. Larson, & H. Preskill (Eds.), *New directions in program evaluation: Organizations in transition: Opportunities and challenges for evaluation. 49* (pp. 5–15). San Francisco: Jossey-Bass Inc.

Preskill, H. (1994). Evaluation's role in enhancing organizational learning. *Evaluation and Program Planning, 17*(3), 291–297.

Preskill, H. & Torres, R. T. (1996, November). *From evaluation to evaluative inquiry for organizational learning.* Paper presented at the annual meeting of the American Evaluation Association, Atlanta.

Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation Practice, 17*(3), 151–161.

Shadish, W. R., & Epstein, R. A. (1987). Patterns of program evaluation practice among members of the Evaluation Research Society and Evaluation Network. *Evaluation Review, 11*(5), 555–590.

Shulha, L. M. (1994, November). *Learning to participate in participatory evaluation: A case study in the transformation of design.* Paper presented at the annual meeting of the American Evaluation Association, Boston, MA.

Shulha, L. M. & Wilson, R. J. (1995). Inviting collaboration: Insights into researcher–school community partnerships. In J. B. Cousins & L. M. Earl (Eds.), *Participatory evaluation in education: Studies in evaluation use and organizational learning* (pp. 115–139). London: Falmer Press.

Smith, N. L., & Chircop, S. (1989). The Weiss-Patton debate: Illumination of the fundamental concerns. *Evaluation Practice*, 5–13.

Stahler, G. J. (1995). Improving the quality of federal human services national demonstration programs. *Evaluation and Program Planning, 18*(2), 129–141.

Stevens, C. J., Dial, M. (1994). What constitutes misuse? In Stevens & Dial (Eds.), *New directions for program evaluation: Guiding principles for evaluators. 64,* (pp. 3–14). San Francisco: Jossey-Bass.

Thompson, B. (1994, April). *The revised program evaluation standards and their correlation with the evaluation use literature*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 370 999)

Thompson, B. & King, J. A. (1981, April). *Evaluation utilization: A literature review and research agenda.* Paper presented at the annual meeting of the American Educational Research Association Los Angeles CA: National Institute of Education. (ERIC Document Reproduction Service No. ED 199 271)

Timar, T. (1994). Federal education policy and practice: Building organizational capacity through Chapter 1. *Educational Evaluation and Policy Analysis, 16*(1), 51–66.

Tomlinson, C., Bland, L., & Moon, T. (1993). Evaluation utilization: A review of the literature with implications for gifted education. *Journal for the Education of the Gifted, 16*(2), 171–189.

Torres, R. T., Preskill, H. S., & Piontek, M. E. (1996). *Evaluation strategies for communicating and reporting: Enhancing learning in organizations.* Thousand Oaks, CA: Sage.

Vroom, P. I., Colombo, M., & Nahan, N. (1994). Confronting ideology and self-interest: Avoiding misuse of evaluation. In C.J. Stevens & M. Dial (Eds.), *New directions for program evaluation: Preventing the misuse of evaluation. 64,* (pp. 49–60). San Francisco: Jossey-Bass.

Weiss, C. H. (1973). Where politics and evaluation research meet. *Evaluation, 1*(3), 37–45.

Weiss, C. H. (1977). (Ed.). *Using social research in public policy making* New York: D. C. Heath and Company.

Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review, 39,* 426–431.

Weiss, C.H. (1988a). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice, 9*(1), 5–19.

Weiss, C. H. (1988b). If program decisions hinged only on information: A response to Patton. *Evaluation Practice, 9*(3), 15–28.