Researching Evaluation Influence: A Review of the Literature

James Leslie Herbert[1]

Abstract

Background: The impact of an evaluation is an important consideration in designing and carrying out evaluations. *Evaluation influence* is a way of thinking about the effect that an evaluation can have in the broadest possible terms, which its proponents argue will lead to a systematic body of evidence about influential evaluation practices. **Method:** This literature review sets out to address three research questions: How have researchers defined evaluation influence; how is this reflected in the research; and what does the research suggest about the utility of evaluation influence as a conceptual framework. Drawing on studies that had cited one of the key evaluation influence articles and conducted original research on some aspect of influence this article reviewed the current state of the literature toward the goal of developing a body of evidence about how to practice influential evaluation. **Results:** Twenty-eight studies were found that have drawn on evaluation influence, which were categorized into (a) descriptive studies, (b) analytical studies, and (c) hypothesis testing. **Conclusion:** Despite the prominence of evaluation influence in the literature, there is slow progress

[1] University of South Australia, Adelaide, South Australia, Australia

Corresponding Author:
James Leslie Herbert, University of South Australia, GPO Box 2471, Adelaide, South Australia 5001, Australia.
Email: james.herbert@unisa.edu.au

toward a persuasive body of literature. Many of the studies reviewed offered vague and inconsistent definitions and have applied influence in an unspecified way in the research. It is hoped that this article will stimulate interest in the systematic study of influence mechanisms, leading to improvements in the potential for evaluation to affect positive social change.

Introduction

While the use of research is important, the success or quality of research does not necessarily depend on its use or impact. However, evaluations that are not used, regardless of their quality, tend to be considered failures (Grob, 2003; Patton, 1997); the job of an evaluator is often not just to produce findings, but to practice in a way that is likely to have an impact (Patton, 2008), and also to go some way toward fostering the implementation of findings (Lawrenz, Gullickson, & Toal, 2007). The development of an understanding of *evaluation use* has been the subject of significant research and theoretical interest since the 1970s, and has been both responsive to and influential on the modern practice of evaluation. From early decision-based models, conceptualizations of evaluation use have developed to include more subtle impacts and attention to the effect that the process of an evaluation can have. More recently, *evaluation influence* has been suggested as another way to think about the effects of an evaluation. This introductory section reviews the development of research into the impact of evaluation over a broad time period, providing some context for the main thrust of the article, the review of evaluation influence.

Use/Utilization: 1970–1986

Early research tended to conceptualize evaluation use in terms of its direct impact on a decision, in some studies this meant simply examining whether or not recommendations of the evaluation had been followed (e.g., Caplan, 1976; Heldt, Braskamp, & Filbeck, 1973). Many of these early studies have been criticized as relying on a flawed standard of evidence, particularly the uncritical use of self-report measures and the lack of triangulation or other

means of verifying behavioral changes (Leviton, 2003). Reflecting on the definition of use in other studies of the era, Alkin, Daillak, and White (1979) were critical of the short reference points used, arguing that the impact of an evaluation may take months or even years to manifest. Studies also construed recommendations not acted upon as examples of nonuse, ignoring that evaluation could inform a decision but not necessarily change it, and that decision makers can often have good reasons to ignore evaluation findings (Birkeland, Murphy-Graham, & Weiss, 2005).

While not used in the way that evaluators expected, many early studies found evaluation findings to have considerable impact. Despite the apparent disregard decision makers had for the recommendations of evaluators in their decisions (e.g., Caplan, Morrison, & Stambaugh, 1975; Patton et al., 1977), researchers found that such decision makers valued research and evaluation information highly (Florio, Behrmann, & Goltz, 1979; Weiss & Bucuvalas, 1977). Resolving this contradiction, researchers found that evaluation findings frequently made important contributions to decision making. First, by influencing the management and practices within programs (Alkin et al., 1979; Becker, Kirkhart, & Doss, 1982); second, by changing the way problems were understood by decision makers (Caplan et al., 1975); and third, researchers recognized evaluation as one of many pieces of information that informed decision makers (Weiss, 1987).

The most enduring taxonomy of evaluation use has been the distinction between instrumental, conceptual, and symbolic/persuasive use (Leviton & Hughes, 1981; Rich, 1977), even in current literature. *Instrumental use* began as a direct, documented, and specific use that researchers expected to observe (Rich, 1977), but has over time come to include the effect an evaluation has over longer periods of time and through a variety of indirect agents (Weiss, Murphy-Graham, & Birkeland, 2005). This type of use depends on evaluation results being the basis of a decision. *Conceptual use* began as the influence evaluation has that cannot be linked to a specific documented use (Rich, 1977). This definition has developed to describe a situation where the use of an evaluation is not direct, but rather the information is absorbed into the common knowledge and comes to form a part of the frame of reference for decision makers (Weiss & Bucuvalas, 1980). This "enlightenment" (Weiss, 1979, p. 429) has been characterized as one of the most important means by which evaluation can assert influence. The *symbolic use* of evaluation involves involvement in an evaluation for ulterior motives or self-interest (Johnson, 1998). Alkin and Taut (2003) make a minor distinction between legitimative use, the use of evaluation to legitimize a previous decision, and symbolic use,

which they describe as conducting evaluation as a symbolic act without intending to use the findings. Similarly, evaluation can be used to delay or avoid making changes, akin to Weiss's (1979) tactical model of social science research utilization.

While certainly conceptual use was understood and recognized in this phase of evaluation research, the Cousins and Leithwood (1986) systematic review served as an important milestone in the development of evaluation research; Cousins and Leithwood (1986) provide a useful summary of the state of research in their review of 64 evaluation use studies. Evaluation results were defined as any information associated with the outcomes of the evaluation, including data, interpretations of data, findings, and recommendations, communicated at any point in the research (Cousins & Leithwood, 1986, p. 332). The types of use found in the studies were categorized in the following way: use as support for discrete decisions; use as education for decision makers; use as constituted by psychological processing; potential use (Cousins & Leithwood, 1986, pp. 341–246).

The studies included in Cousins and Leithwood (1986), while focused on results based use, suggest the future development of the influence framework through categories of studies that recognize more subtle effects than direct use. The review served to consolidate research knowledge in a way that opened up important issues for the evaluation community to deal with.

Evaluation Use/Impact: 1986–2000

According to Shulha and Cousins (1997), the most important development following Cousins and Leithwood (1986) was the increasing importance of context. Evaluation researchers were concerned with how evaluation results interact with other influences in decision-making processes and how evaluators should engage with this process. While contributing to this, Patton (1998) also highlighted the experience of participants involved in evaluation; that change can occur through the process of an evaluation. This observation has in part driven modern approaches to a participatory and collaborative evaluation process. These advancements were influential on practice as well as leading to the next phase of research, where researchers sought to theorize and observe the subtle influences of evaluation findings and processes.

One of the most important events in the development of evaluation theory post-1986 was the Weiss–Patton debates. Each being luminaries in the evaluation field, but encapsulating very different visions for the profession, the exchange polarized and energized debate and theory

on the ideal nature of evaluation use. Weiss (1988b) argued that expecting evaluations to routinely result in instrumental use is unrealistic. She advocated evaluators accepting the reality of organzational decision making and to focus on producing sound evaluations, encouraging evaluators to aspire to instrumental use, but not to use *instrumental use* as sign of the success or quality of an evaluation. Patton's (1988) position was that evaluators should actively engage in fostering the usefulness of evaluation by delivering the information and processes that meet the needs of the commissioners of the evaluation. Patton (1986) had already outlined this position in his book *Utilization-Focused Evaluation*, which served as a manual for evaluators engaged in the type of collaborative and consultative processes required for his vision of evaluation. Shulha and Cousins (1997) link the Weiss–Patton exchange to increased theory and research on the nature of context and evaluation use, particularly in work on evaluation epistemology, the political frame of reference in decision making, psychosocial processing of evaluation information, and organizational culture and learning.

Equally influential on the theory and practice of evaluation around this era was the notion of process use being a mode of utilization; that the experience of participation in an evaluation can be as influential on behavior as any recommendations or reports (Patton, 1997). Process use has been discussed as a type of use, alongside instrumental and conceptual use (Sandison, 2006), leading to collaborative and participative approaches that aspire to equitable power relationships, organizational learning, and stakeholder empowerment (Cousins, 1996; Cousins & Earl, 1995). Process use remains an important mode of use, reflected in recent models of evaluation influence (Henry & Mark, 2003a; Kirkhart, 2000; Mark & Henry, 2004).

Evidence of a rethinking of evaluation in terms of context and the source of influence is most evident in the practices of evaluators and the intended uses for which funders and programs employed evaluation. Evaluators were seen to aspire to collaborative and egalitarian relationships with stakeholders (Cousins & Earl, 1995), to foster the skills and qualities of evaluation within organizations (Preskill, 1994), and to actively promote the use of evaluation among intended users (Patton, 1997). Evaluators and the commissioners of evaluations recognized a variety of uses or benefits that evaluation could produce depending on the needs of the program. Through process use and the study of evaluation context, participant learning became an important goal of evaluation, alongside informing decision making.

Evaluation Influence: 2000 to Present

Around the year 2000, the evaluation community began discussing and debating frameworks of evaluation influence, a possible successor to theories of use. In many ways, this approach reflected Weiss' (1988a) vision for evaluation, while incorporating the changes in practice and understanding that Patton (1997, 1998) had brought to the field. Critical of past definitions of use, a body of theory and research has sprung up in a relatively short time to provide a more nuanced understanding of the influence of evaluation, while acknowledging the importance of the past 30 years of research. Encapsulating existing insights and approaches, evaluation influence offers a comprehensive framework with which to consider the intended and unintended impacts that evaluation can have, which is particularly important considering the more indirect goals of modern evaluation (e.g., organizational learning and empowerment) and an onus on evaluators to contribute to better social conditions through their work (Henry, 2000).

While evaluators have long expressed dissatisfaction with the definition of evaluation use (e.g., Alkin et al., 1979; Caplan, 1980), this became the focus of discussion around the turn of the century. Kirkhart (1995, 2000) called the scope and language of use awkward, inadequate, and limiting. Henry (2000) suggests that use has been embraced as the "holy grail of evaluation" (p. 85), which he presents as an unworthy, unhelpful, and self-serving goal that may limit the contribution of evaluation to improving social conditions. Proponents of evaluation influence argue the central problem has been the vagueness of the term, with inconsistent definitions existing in the literature (Henry & Mark, 2003b; Kirkhart, 2000; Mark, 2008). Evaluation research depends on clear and consistent definitions and language in order to make sense of what is known (Patton, 2000); Henry and Mark (2003a) and Kirkhart (2000) argue that the term *use* tends to underestimate the impact of evaluation by emphasizing results based use and that *use* suggests an intentionality, immediacy, and directness that may not always exist. Mark and Henry (2004, p. 37) suggest that "contemporary theories of use (or evaluation utilization) are simultaneously impoverished and overgrown." Although researchers have attempted to update the concept in line with new understandings about evaluation use, these ad hoc additional elements lack a clear framework (Christie, 2007; Kirkhart, 2000; Mark & Henry, 2004; Weiss et al., 2005). Influence has been suggested as a remedy, either as an extension of use (Alkin & Taut, 2003; Kirkhart, 2000) or as a replacement for it (Henry & Mark, 2003a; Mark & Henry, 2004). Influence, by contrast, represents an approach to studying the effects of evaluation in

the broadest possible terms, across the indirect, unintended, and long term (Kirkhart, 2000).

By shifting from *use* to influence, proponents have attempted to legitimize the study of the full impact of evaluation (Kirkhart, 2000). Influence—which according to Henry and Mark (2003a) includes all " ... *evaluation consequences that could plausibly lead toward or away from social betterment*" (p. 295)—adds to the scope of impact, not to obscure use, but in order to understand the mechanisms and processes that may be antecedents of use. Mark (2011) suggests that evaluation influence represents an attempt to tell the whole story of change. Shifting to a framework of evaluation influence is argued to have the following benefits (Kirkhart, 1995, 2000; Mark, 2003; Mark & Henry, 2004):

- influence provides a definition and a framework that reflects the full impact of evaluation and a cohesive way to organize theoretical and empirical knowledge of the effect evaluation can have on programs;
- by adopting this more comprehensive view, influence allows for the study of implicit mechanisms that affect change, including processes at the individual, interpersonal, and collective levels;
- influence frameworks are oriented around linkages to more developed constructs in other fields of literature such as attitude change, priming, skill acquisition, and persuasion;
- shifting to an influence framework allows for the study of pathways of influence and the study of situations where evaluation failed to affect change; and
- influence is built around social betterment as the ultimate goal of evaluation, rather than use.

Other researchers have argued that influence is better suited to the study of the effects of evaluation in an organizational context (Cousins, 2004; Poth, 2008; Weiss et al., 2005), drawing on well-established mechanisms of change from the social sciences (Leviton, 2003; Weiss et al., 2005). The development of an integrated framework of influence has also been argued as an effective way of acknowledging the broad effects of an evaluation (Caracelli, 2000; Weiss et al., 2005), in a way that Leviton (2003) suggests is useful for developing evaluation practice. Moreover, this shift is argued to be vital for the evaluation research community, allowing for a more detailed framework that will enhance the evidence base for evaluation practice (Mark, 2011; Mark & Henry, 2004), which Mark (2001, 2008) has criticized as being overwhelmingly expert-based and susceptible to fads and ideology.

The suggestion of a shift to influence as the conceptual framework with which to study the effects of evaluation has not been without its critics, who argue that replacing use is unhelpful in informing evaluation practice (Alkin & Taut, 2003; Hofstetter & Alkin, 2003; Patton, 2008). Alkin and Taut (2003) suggest that the influence concept is not helpful, as it includes events and factors outside the awareness or control of an evaluator. Social betterment, an important part of Mark and Henry's model (2004), is also criticized as being unrealistic and impractical (Cousins, 2004; Patton, 2008). Hofstetter and Alkin (2003) and Patton (2008) offer program improvement as the purpose of evaluation and the prerequisite for any social betterment than can follow. The Henry and Mark (2003a) conceptual framework of influence has been criticized by McEathron (2008) as a "hodgepodge of unparallel processes" (p. 42) that do nothing to resolve the problems with evaluation use. She concedes that this has been addressed by Mark and Henry's (2004) framework and the inclusion of levels of analysis (e.g., cognitive/affective, motivational, and behavioral), but is critical of the loss of focus on the individual/interpersonal/collective levels, and of a "rational, linear, uni-directional, pro-innovation adoption or use of the evaluation" (McEathron, 2008, p. 43). These criticisms seem token considering Mark and Henry's (2004) framework of mechanisms are still organized around the individual/interpersonal/collective levels, and that part of analyzing the pathways of influence is about attention to more subtle processes that may not result in policy or practice change.

While the relative merits of conceptualizing the impact of evaluations in this way remains under debate, evaluation influence represents a developed and nuanced framework to build a body of research around. Despite a long history of research in evaluation use, the field of evaluation is no closer to evidence based practice in terms of how to affect social change (Mark & Henry, 2004). Evaluation influence, in the form proposed by Mark & Henry (2004), suggests that building a body of evidence begins with recognizing the basic mechanisms of influence that can accrete and result in program level change. Through this, researchers can develop knowledge about the factors important to different types of influence, and evaluators can adapt their practice to emphasize different types of influence. The main criticisms of the approach are that it includes factors that can't be controlled or foreseen by evaluators (Alkin & Taut, 2003). While certainly some of the factors related to influence may be out of control of evaluators, there is still value in understanding these factors and in developing a body of literature that recognizes some of the difficult circumstances evaluators practice in.

The Evaluation Influence Frameworks

Implicit in the change to influence are the conceptual frameworks that provide some substance to what would otherwise be a point of semantics. Figure 1 presents these frameworks in order to contrast the difference between Kirkhart (2000), Alkin and Taut (2003), Henry and Mark (2003a). Consistent with their argument for influence to be understood as an extension of use, Kirkhart (2000) and Alkin and Taut's (2003) models incorporate *use*, with awareness as the threshold for influence. Henry and Mark (2003a) and Mark and Henry's (2004) frameworks construct influence as the broadest understanding of the effect an evaluation can have (Table 1), subsuming evaluation use within it, and advocating for thinking about and researching influence in terms of underlying mechanisms.

Kirkhart's (2000) framework suggests three dimensions: Intention is simply if the evaluator intended the influence; source reflects the idea that the process and the findings of an evaluation can have influence; and time which is split into immediate, when the evaluation is concluded, or longer term. Alkin and Taut's (2003) framework slightly differs with the inclusion of awareness, which they suggest represents the difference between use and influence.

Over two articles, Henry and Mark (2003a) and Mark and Henry (2004) have refined a more sophisticated framework, building on Kirkhart's (2000) work, but emphasizing the mechanisms through which evaluation can be said to affect change. Henry and Mark (2003a) present influence as interconnected change mechanisms at the individual, interpersonal, and collective levels. As displayed in Table 1 Mark and Henry (2004) elaborate on this model, categorizing influence into families of similar mechanisms (general, cognitive and affective, motivational, and behavioral). Mark and Henry (2012) have since also suggested a family of relational mechanisms embedded in "aspects of ongoing relationships, structures and organizational processes" (Mark, 2011, p. 115). A key idea in these frameworks is the interconnection of influence as a kind of chain reaction of events. This framework aims for a comprehensive understanding of the complex, contextual, and often convoluted series of processes that can lead to change and just as often lead to no observable change.

An important feature of Mark and Henry's (2004) framework is the linkage to more established constructs in research and theory in psychology, political science, organizational behavior, and sociology. Drawing on knowledge and terminology from other disciplines allows for the enrichment and integration of evaluation research with parallel areas of inquiry. Mark and Henry (2004) also argue that the influence framework will
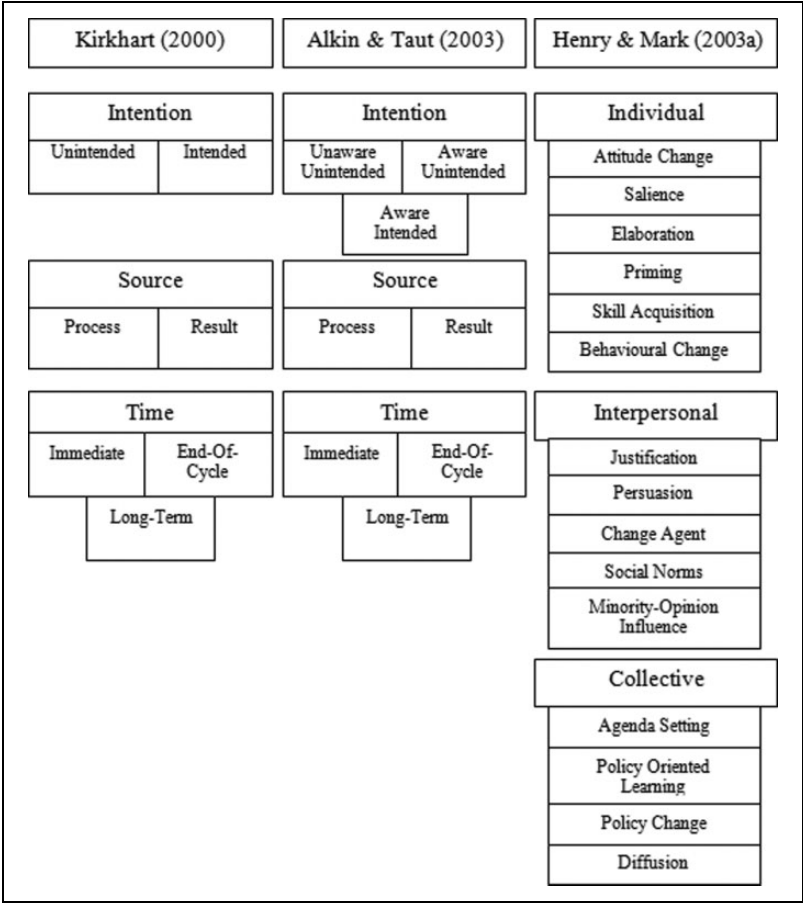
Figure 1. A comparison of evaluation influence frameworks (adapted from Alkin & Taut, 2003; Henry & Mark, 2003a; Kirkhart, 2000).

potentially serve as a stimulus for systematic inquiry to inform evaluation practice as an antidote to "expert-based ... evaluation practice" (Mark, 2008, p. 114). They suggest that evaluation influence represents an opportunity to develop a structure of issues for research involving simple questions that can be built into a collective body of knowledge.

Reflecting the developed state of the area, researchers have recently sought to consolidate the Mark and Henry (2004) model. Fleming (2011)

Table 1. An Expanded Typology of Evaluation Influence Mechanisms (From Mark & Henry, 2004, p. 41).

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

presents the social psychology research underpinning some of the evaluation influence mechanisms, assessing the quality of some of the suggestions in the evaluation literature on increasing influence. Appleton-Dyer, Clinton, Carswell, and McNeill (2012) propose a conceptual model for evaluation influence applied to partnership arrangements in the public sector, providing a sophisticated approach to the analysis of influence in the complex interactions between organizations. Mark (2011) provides some suggestions about researching evaluation influence. Overall, he suggests a diversity of research methods especially given that influence research is in its infancy, but favors an increase in more directed and narrow research that explores specific influence mechanisms and chains of influence.

While still a maturing area of research and theory, evaluation influence has been prominent in discussions about the impact of evaluation. Given recent interest in directing research on evaluation influence to produce a critical mass of evidence to inform evaluation practice (Lawrenz, King, & Ooms, 2011; Mark, 2008, 2011) and in the context of enduring criticisms of the approach (Alkin & Taut, 2003; Hofstetter & Alkin, 2003; McEathron,

2008; Patton, 2008), it seems timely to explore the state of the literature on evaluation influence. Despite its prominence in the literature, and the slow accretion of empirical research on the topic (Gildemyn, 2014), there has not yet been a comprehensive review of evaluation influence research.

A Review of Evaluation Influence

This review sets out with the aim of exploring the current state of evaluation influence in the literature, particularly the extent to which influence has been used as the conceptual basis for research on the impact of evaluations. The envisioned purpose of this was to support the integration of the literature on evaluation influence, to provide a resource showing how evaluation influence has been operationalized in research, and examine the utility of drawing on evaluation influence as part of the conceptual foundation of research, meta-evaluation, or in planning the theory of change for an evaluation. The review sets out with the following research questions:

- How have studies that have cited evaluation influence defined influence in their research;
- How is this definition of evaluation influence reflected in the way the research was conducted; and
- What do the findings of these studies suggest about the utility of evaluation influence as a conceptual framework in which to consider the impact of evaluations?

The review was initially conducted in July 2011, drawing on all articles that cited Henry and Mark (2003a), Kirkhart, (2000), and Mark and Henry (2004) and conducted original research drawing on evaluation influence in some form. Articles were screened based on a reading of the abstracts; the reviewer then obtained articles that presented original research about evaluation, in order to identify studies that drew on evaluation influence as part of the conceptual foundation of the research. For articles that fit this criterion, the term influence was searched in the text of each of the articles in order to quickly establish how the authors had used the concept in their research. Studies that had researched some aspect of evaluation influence were then read in their entirety and summarized into an annotated review document. Some articles that appeared relevant—mainly PhD theses—were not able to be obtained in full text. A number of articles had English language abstracts, with the main text in another language; these were also not incorporated into the review. The process of searching for eligible studies

was repeated in April 2013, focusing on articles published since the initial review. Just prior to submitting the article, the reviewer was made aware of the recently published study by an e-mail update from a journal. This study (Gildemyn, 2014) was included in the review as one of the few empirical investigations of evaluation influence.

While widely cited and discussed, relatively few studies in the literature directly explore evaluation influence with empirical research. Indeed, a number of researchers have remarked on the lack of empirical support and the lack of take-up of the approach (Murphy, 2007; Poth, 2008), although considering the paucity of empirical research into evaluation generally (Henry & Mark, 2003b) the topic is quite well researched. Twenty-eight studies were found that sought to research aspects of evaluation influence, employing a variety of definitions, and methods, differing considerably in how influence was incorporated into the research. These studies are presented as categories: (a) *descriptive research* where influence concepts were used to describe how the impact of research played out; (b) *analytical research*, which uses influence concepts to explain or infer how examples of influence came about; and (c) *hypothesis testing* that directly tests hypotheses about the mechanisms underlying influence in a real world or laboratory setting. It should be noted that the vast majority of the articles originally reviewed cited evaluation influence as a mere footnote to their discussion of the evaluation use literature (e.g., Garcia, 2008), or in order to critique it to support their preference for evaluation use (e.g., Murphy, 2007). While these studies contribute to knowledge about the impact of evaluation, the scope of this review is tightly on studies that directly drew on evaluation influence as part of the conceptual basis for their research in order to enhance understanding about how influence has been defined in research, and what this research suggests about the viability of evaluation influence as the conceptual basis for research, or as a consideration in planning evaluations.

Descriptive Research

Nine of the studies included in this review (32%) are descriptive, in that evaluation influence was used as a framework to present how the influence of an evaluation played out over time. Much of this research presents as a kind of meta-evaluation, or as examples of types of mechanisms in the context of a discussion article about evaluation influence. While illustrative of mechanisms, and the interconnection of influences that is central to the concept of evaluation influence (Mark, 2011), these studies lack any systematic analysis of why influence did or did not occur, and have a limited basis to

suggest approaches to fostering influence. This is not to disparage these studies, many of which have explored evaluation influence in the context of real evaluations in order to demonstrate the value of considering the broader range of impacts an evaluation can have, but to be clear about the limitations of these studies in terms of explaining *why* influence occurred as opposed to *how*.

Some of the descriptive studies included in the review have drawn on the Kirkhart (2000) framework in their description of the impacts of evaluation, interpreting influence as attentiveness to the unintended effects of an evaluation over a longer timescale. Benjamin and Misra (2006) undertook 13 interviews with staff within agencies that funded nonprofit services to examine the influence of performance accountability across the sector. The research used the Kirkhart (2000) framework to structure the findings, describing how participants had experienced performance accountability in their organizations, framing performance accountability as a form of evaluation. Rebolloso, Baltasar, and Canton (2005) also drew on Kirkhart's (2000) interpretation of influence in their investigation of the influence of two public education evaluations, comparing a capacity building approach to a more traditional evaluation design. The researchers highlight the value of attending to influence through a variety of diffuse effects including an improvement in participant attitudes about evaluation and understanding the perspectives of other people in the organization.

As with the studies that adopted the Kirkhart (2000) framework as a way to broaden their scope of effects, some studies have simply used some of the broad categorizations in evaluation influence as a way to frame their investigation of the impacts of evaluations. Cowley and Good (2010) examined the influence of an evaluation on education staff attitudes and behavior related to their technical assistance work. Henry and Mark's (2003a) influence mechanisms are used to frame examples of how the evaluation influenced change across the individual/interpersonal/collective levels. Lawrenz, King, and Ooms (2011), with an awareness of evaluation influence, describe the relationship between involvement in a multisite evaluation and the use of that evaluation, focusing on the "use of evaluation by secondary, somewhat unintended users" (p. 50). Looking at four multisite evaluations, the researchers detail how greater involvement led to more instances of unintended use, which then translated into further instances of use. In a brief and informal description of three evaluations, Henry (2003) illustrates the kinds of situations where evaluation may be influential, using the cases as exemplars, providing suggestions for practice illustrated by examples, rather than arrived at through

any analysis. Henry (2003) presents the importance of illustrating the monetized social benefit compared to the cost of the type of intervention the program was based on, improving practices within the program, and changing perceptions about the effectiveness of a program. He suggests that good evaluation studies can set the agenda in terms of defining key variables and the instruments used to measure them, this often depends on their technical quality.

Fewer studies went to the lengths of attempting to identify examples of evaluation influence mechanisms in describing how the impact of an evaluation played out. Weiss, Murphy-Graham, and Birkeland (2005) reported on a long-term case study on the high-profile Drug Abuse Resistance Education program in the United States, and attempted to undertake an in-depth analysis of influence mechanisms to complement their study of evaluation use. They reported that efforts to identify the pathways of influence were challenging, partly due to the time gap between the conclusion of the evaluation and their fieldwork. It should be noted that their interview protocol was not specifically designed around efforts to identify influence mechanisms. Vataja (2011) also explored both use and influence in looking at learning and development among eight internal improvement-oriented evaluations that used either empowerment evaluation approaches or a model described as the ITE method (short form for a Finnish term *participant themselves*). Each of the approaches encouraged participants to employ evaluative thinking in their approach to work, with examples of their influence described. Mark and Henry's (2004) evaluation influence mechanisms are drawn on in describing how the process of undertaking the evaluations resulted in change. Diaz-Puente, Yague, and Afonso (2008), and Diaz-Puente, Montero, and Carmenado (2009) examined how influence played out over a 10-year period in a series of evaluation capacity building projects in rural Spain. As a series of case studies, the researchers have used the research as an opportunity to present the change processes that occurred over the course of conducting a sustained empowerment evaluation project. Changes in attitudes and actions at the individual, interpersonal, and collective levels were discussed, along with some description of change mechanisms (e.g., capacity building and salience) linked to the evaluation findings and process. These studies also went to some effort to study the pathways of influence, tracing the interaction between different mechanisms. While both studies are primarily descriptive, Diaz-Puente et al. (2009) do suggest some critical components to successful (influential) empowerment evaluation and rural development informed by their analysis of the evaluation.

Analytical Research

Most of the studies ($n = 15$; 54%) involved research that has attempted to analyze evaluation influence to attempt to explain why (or why not) influence has occurred. While many of these are case studies similar to the descriptive category, these studies are distinct as they have observed or attempted to reconstruct evaluations for the purpose of being able to make some inferences about factors associated with evaluation influence. There is an inherent challenge with this type of retrospective research, particularly when they take place long after the evaluation has concluded. These studies are primarily explorative, involving broad primarily qualitative approaches.

As with the descriptive category, some of the studies have adopted the Kirkhart (2000) model of evaluation influence, meaning that the focus is on factors that foster influence over time, across both the evaluation process and results, and both inside and outside of the evaluator's awareness. Alexander (2003) presents an analysis of the influence of health service evaluations on practice, drawing on three evaluations as case studies to present how influence played out, and some of the moderators of influence. Kirkhart's (2000) model of evaluation influence is drawn on, with cases selected based on an existing typology (Dahler-Larsen, 2001) of the relationship between evidence and practice in health service evaluations (tragic, magic, and competing). The analysis comparing the cases suggests that the key differences in influence seem to be mediated by the successful collaboration between evaluation staff and staff involved in the program. While the study identified many other differences (e.g., a focus on service users and lack of funding to implement findings), there is little that can be said about how all these factors might interact, or indeed what the most important factors might be. Morabito (2002) presents a brief case study on increasing the influence of the evaluation process through the use of distinct roles for evaluators. Drawing on Kirkhart (2000), the researcher has set out to identify "any organization-related change stimulated during the evaluation process" (Morabito, 2002, p. 322). The article focuses on a categorization of roles played by the evaluator in the case study and how these may enhance the influence of the evaluation process, connecting the performance of these roles to particular types of influence in the case study. Poth (2008) cites Kirkhart (2000) as guiding her approach to be attentive to a range of impacts in her case study research on how stakeholder engagement results in the influence of evaluation. The researcher's own experience of stakeholder engagement is critically analyzed against a set of engagement principles, which are elaborated and expanded on using the case study.

Weets (2008) also applies Kirkhart's (2000), arguing that the measurement of performance audits typically fails to consider the importance of influence (i.e., longer time scales, unintentional influence). The researcher presents a series of case studies on the effectiveness (influence) of performance audits, which are analyzed with a combination of approaches (auditees' perceptions, impacts, and contribution to public debate) and are used to suggest some factors associated with effectiveness.

Some studies set out to examine the factors related to influence drawing on Mark and Henry (2004), but without seeking to identify and understand the circumstances around evaluation influence mechanisms. Eschewing the language of influence, and indeed offering significant critique of the evaluation influence concept, McEathron (2008) concludes her study of independent science review panels by considering the implications of her research for evaluation practice. Three case studies of the scientific basis for decision making in natural resource management were used to develop a set of salient characteristics at the individual, interpersonal, and collective levels, which McEathron (2008) suggests may have relevance to evaluation practice. Incorporating Valovirta's (2002) observations about argumentation into the interpersonal level of influence, Lehtonen (2010) built on Henry and Mark's (2003a) framework to explore influence through interviews, a document analysis and a stakeholder workshop. While setting out to analyze the influence of policy performance indicators, the approach also examined the factors relevant to the influence of the indicators across the U.K. energy system. The analysis of the case study suggests that the direct use of expert knowledge and information in policymaking is rare; however, the indicators resulted in many indirect impact identified by attending to influence. Lehtonen (2010) outlined that the interviewees emphasized the importance of the reliability, validity, and timeliness of the indicators, but that the analysis suggests the lack of influence is explained by the interaction of user and policy factors.

In a methodologically distinct line of research, Greenseid, Johnson, and Lawrenz (2008) and Greenseid (2008) used citation analysis as a means of measuring the influence of particular evaluations, while also exploring the types of factors that seem to have an impact on influence. Over a series of studies, the researchers have drawn on influence as the conceptual foundation for their work, although without connecting their definition to any particular evaluation influence mechanisms. The research was in part undertaken in order to explore the validity of using citation analysis to measure influence. While suggesting that citation analysis is a useful way of measuring influence, Greenseid (2008) acknowledges the limitations of the

approach, pointing to the importance of understanding the content of citations, and of other means of assessing influence. Greenseid et al. (2008) suggest that their citation analysis supports the idea that large and collaborative evaluations tend to be highly cited, in particular by the staff involved in the evaluation. Roseland, Greenseid, Volkov, and Lawrenz (2011) presented an analysis incorporating citation analysis, an online survey, and surveys and interviews with senior program staff. Despite the citation analysis suggesting each of the evaluations they studied were quite influential, surveys and interviews suggested that program staff had very limited knowledge of the evaluations and found it difficult to identify the influence of the evaluations. Similarly, Greenseid and Lawrenz (2011) used a citation analysis of four large multisite program evaluations to find that evaluation instruments and reports tend to be the most highly cited products of an evaluation. The researchers also suggest that other factors such as the reputation of the evaluation team and the uniqueness of the research may also play a part in the level of influence.

Of significant interest in this review are studies that drew on the evaluation influence mechanisms (Henry & Mark, 2003a; Mark & Henry, 2004) in research order to make some inferences about why particular types of influence occurred. Burr (2009) drew on both evaluation use and influence in developing a survey examining the effect that evaluation had on project directors of 17 university preparation programs. The survey items reflect the levels of influence and the change mechanisms of Henry and Mark (2003a), alongside efforts to examine the instrumental, conceptual, symbolic, and process use of evaluation. The researcher reported influence occurred through all of Henry and Mark's (2003a) change mechanisms, with each of them being incorporated into a survey instrument reporting on the influence of the evaluation. It should be noted that the items on this instrument are the same items used to measure evaluation use, for example, ''learn about the weakness of my program'' measures both attitude change and conceptual use. While certainly use and influence have significant overlap, influence is presented as a recategorization of the same impacts, which in many ways defeats the purpose of evaluation influence as laid out by Mark and Henry (2004). The research also draws on the ratings of participants to rank Cousins and Leithwood's (1986) factors impacting the evaluation use, finding the relevance of the evaluation to program directors, and their commitment to evaluation being the most importance factors for use.

Of most interest in this section are studies that not only incorporated specific influence mechanisms into the analysis in order to understand why specific types of influence occurred, but set out to explore their interaction.

Using Mark and Henry's (2004) framework, Cheng (2006) undertook a case study of the influence of two evaluations of literacy programs, using retrospective, semistructured interviews and document review to assess the influence of program evaluations on literacy instruction. The researcher outlined the influence mechanisms observed in the research and reported some success in tracing some of the pathways of influence that followed from the evaluation. Cheng attempted to use the framework but found great difficulty, as the changes that resulted from the evaluation were connected to multiple change mechanisms, with great difficulty in identifying the step-by-step pathways. The analysis resulted in the identification of three factors that appeared to be related to use or influence: human factors, structures/resources, and external/contextual factors.

Fjellström (2007) used the Mark and Henry (2004) framework in describing the influence of a collaborative evaluation of a teaching initiative, suggesting that the " ... analysis model strongly contributed to the rich description of evaluation effects" (p. 29). Presenting an analysis of influence including attempting to present the chains of influence, the researcher suggests that the analysis supports the importance of ownership and deliberation in the influence of deliberative evaluation.

Oliver (2008) undertook a multiple case study of international nongovernmental organizations' (NGOs) emergency responses; tracing the pathways of influence from evaluations of these responses and the influence these had in the organization in future emergencies. The research drew on evaluation reports and interviews with people associated with the evaluation. Oliver (2008) goes to some length in attempting to operationalize Henry and Mark's (2003a) mechanisms of influence, developing a detailed checklist and set of definitions associated with each. In developing this criterion, Oliver (2008) observed that mechanisms such as attitude change are reasonably easy to detect as it easily lends itself to a program, while others such as salience or elaboration will be more difficult to discover as they are more associated with policy-related issues, and will be more difficult to pinpoint with interview data. Moreover, Oliver suggests individual level mechanisms will tend to be emphasized where the evaluation and case study and undertaken concurrently. Observing the chain of influence across the evaluations studied, Oliver suggests evaluations often fell short of influence at the collective level because the individuals responsible for setting policy agendas are removed from the process of the evaluation. Also important factors in the interruption of influence chains were the absence of a culture of learning, a lack of institutional memory, the lack of opportunity for staff

to read past reports when a new crisis emerges, and that emergency response was sometimes just a small part of a person's job.

Recently, Gildemyn (2014) presented a study of the influence of civil society organizations' monitoring and evaluation of government programs in Ghana, producing a comprehensive map of evaluation influence mechanisms. The study aimed to explore what mechanisms of influence were employed, while also examining how interface meetings impacted on how influence played out. While not a panacea for influence, Gildemyn found that, particularly in the context of Ghana, the interface meetings were important opportunities for exchange and debate that created an environment for influence mechanisms to occur. The study acknowledges its place as one of the few direct empirical applications of Mark and Henry's (2004) framework (although this review has identified a few others) and suggests a number of additional influence mechanisms discovered in the case study; pledge to action, and onetime action. The researcher also highlights the challenges of getting the timeline right for tracing the pathways of influence, the "trade-off between waiting long enough until sufficient time has elapsed for such changes to be observed, but not too long in order that memory/recall biases related to earlier mechanisms may be reduced" (Gildemyn, 2014, p. 15).

Hypothesis Testing Research

Relatively few studies in this field ($n = 4$; 14%) set out with a hypothesis to test, and for those that did, the research was often employed influence more as a means to test some other hypothesis. Studies included in this section identified evaluation influence as the conceptual foundation of the research, made some specific inference based on existing research and tested their hypothesis in the context of an evaluation or among participants with experience of evaluations.

While not specifically researching evaluation influence, Baptiste (2010) draws on Kirkhart's (2000) conceptualization of evaluation influence in the development of a set of statements aimed to test (a) the extent to which a sample of professional evaluators agreed with a definition of process use and (b) the type of process use that emerges in particular contexts. While the term evaluation use is preferred by the researcher (possibly due to its familiarity with participants), Kirkhart's dimensions of time and intentionality have informed the types of process use statements presented to the participants. The research found that while some evaluators agreed with the standard definition of process use, three other definitions were evident in

the data, and that different contexts can affect the types of process use that occurs. Ledermann (2012), while adopting the language of evaluation use suggests that he is engaged in investigating a small number of Mark and Henry's (2004) influence mechanisms, influenced by Valovirta's (2002) observations about evaluation use as argumentation. The researcher hypothesizes about four different mechanisms or roles for evaluation to play (awakener, trigger, referee, and conciliator) and hypothesizes the types of conditions necessary (i.e., pressure for change, level of conflict, novelty value, and evaluation quality). From 11 program evaluations, Ledermann (2012) conducts a qualitative comparative analysis in order to test out his hypothesized conditions for different types of influence, more or less supporting his typology, but with a few additional mechanisms (endorser and reviser). Evaluation influence is more or less a backdrop to this research, the similarity of the mechanisms explored by Ledermann (2012) resemble many of those found in the interpersonal level of change in Mark and Henry's (2004) framework, and such has some value in terms of adding to the limited body of evidence about the conditions for particular influence mechanisms.

Frey and Widmer (2011) developed a scale of influence to explore a set of hypotheses about the influence of systematic evidence on reviews of Swiss government policy. While adopting the language of influence, and discussing Mark and Henry (2004), influence is defined quite simply as "(the) extent systematic evidence has shaped the contents of the revision process" (Frey & Widmer, 2011, p. 5). From a qualitative analysis of 10 public policy revisions, the researchers found (a) inconclusive evidence about the importance of the availability of efficiency information; (b) policy specialists (e.g., civil servants, NGOs) and members of parliament both highly value effectiveness information, but specialists may value efficiency information more highly; and (c) inconclusive evidence about the use of efficiency information by policy opponents.

Directly addressing the need to develop a body of research on evaluation influence, Christie (2007) undertook a simulation study of how different types of evaluation data influence decision makers at the individual level. This simulation involved nine scenarios with different forms of evaluation evidence and the participants' survey responses on how influential the data were, taking into account the participants' preexisting beliefs in the efficacy of the program described. Christie's research responds directly to Mark's (2008) call for efforts to unpack evaluation influence through simple research questions and to build a body of evidence for the framework and mechanisms of evaluation influence. Participants were likely to be

influenced by all types of data, but the extent of influence differed by educational background, sector of employment, and the degree to which decision makers were informed by other data types. Age, sex, and race did not affect the use of evaluation study data. Where participants had existing beliefs about the effectiveness of particular programs, they were more likely to be influenced by survey data and less likely to be influenced by anecdotal accounts (Christie, 2007). Large-scale evaluation study data influence was highest when participants were asked about implementing a program in their own organization. People with a degree or work background in education were less likely to be influenced by surveys, the researcher explains this as possibly being a backlash to compulsory education testing in the United States. While this approach to testing hypotheses about evaluation influence has value, there are some limitations. While Christie (2007) talks at length about the use of simulated decision-making experiences in similar fields and advisably adds considerable caveats to her findings, there is also the issue that focusing on individual level influence obscures the connection to interpersonal and collective level influence. As an example, how useful is it to know about the decision an individual would make in the absence of organizational context, without the need to conform to institutional structures and norms.

Discussion

This review of the evaluation influence literature sets out with three aims: to get a sense of how researchers have defined influence, how this has translated into the research approach, and what does the current body of evidence suggest about the utility of evaluation influence in research that aims to inform evaluation practice?

How Have Researchers Defined Evaluation Influence?

As stated previously, the review identified many studies that cited influence merely to acknowledge the body of work, or to critique it; this review focused on studies that drew on influence ideas in their research. The review identified a methodologically diverse collection of studies; however, the way evaluation influence is defined seems to fit into one of the following categories:

- *A broader view*: Primarily associated with Kirkhart's (2000) definition of influence, these studies present influence as an attentiveness to the broader effects of an evaluation, particularly in terms of the

longer term impact of an evaluation and in unintended impacts. Influence is seen as an extension of use;

- *Levels of effects*: Considers the broad impact of an evaluation at the individual, interpersonal, and collective levels, with use and influence more or less overlapping in definition, but with influence representing more subtle and delayed effects;
- *A framework of mechanisms*: Looks to identify specific influence mechanisms that parallel types of evaluation use (e.g., conceptual use);
- *Influence pathways*: Identifies specific influence mechanisms, but seeks to connect the chains of influence and identify the interaction between different influence events. Evaluation use is subsumed within evaluation influence.

In part, the definition used reflects the caution of some researchers in favoring this new approach to thinking about the impact of an evaluation, an approach that, as some of the researchers have suggested, can be onerous to pursue. For many of the studies, evaluation influence was a sidebar or a means of addressing a research question of interest, whereas few researchers are addressing questions relating directly to issues with the conceptual understanding of evaluation influence. Recently, Mark (2011) has emphasized that use and influence are not competing concepts, and makes the case for the continued existence of use, but argues for the value of evaluators attending to influence in their practice. Encouraging evaluation practitioners to think about evaluation influence, as opposed to instrumental or conceptual use, encourages a strategic approach to effecting change. Influence has a role as a more academic approach to thinking about the impact of an evaluation, a complex, and confusing tangle of effects and relationships that underlie the simplicity and false certainty of direct and clear evaluation use.

How Has Evaluation Influence Been Researched?

Overwhelmingly, researchers have approached the task of researching evaluation influence with case studies. Researchers either employed retrospective approaches, undertaking interviews, and collecting organizational documents, or presented the events from their perspective as the evaluators, triangulated with interviews. The identification of influence mechanisms was not well explained in many of the studies, with many studies lacking any kind of criteria or process explaining how mechanisms were identified and connected. Some of the difficulties experienced by researchers have to do with the lack of specific procedures suited to identifying influence

mechanisms (e.g., Weiss et al., 2005). In contrast to studies that lacked clear procedures, Oliver (2008) provides a detailed set of operational definitions of the mechanisms and enough detail in the narratives in order to be able to see how the criteria were applied. Gildemyn (2014) also provides significant detail about the families of mechanisms and how they were investigated in her case studies.

Awareness of how the timeline of the research may affect the findings may provide some perspective for findings of case studies that attempt to trace the chains of influence. Studies that allowed a longer interval following an evaluation seem to be more likely to observe the link between these individual changes and collective change (Diaz-Puente, Montero, & Carmenado, 2009; Diaz-Puente, Yague, & Afonso, 2008; Oliver, 2008). Weiss et al. (2005) suggest that in their study the interval had been too long (2–8 years after the events) to adequately capture individual and interpersonal change. Gildemyn (2014) also reflects on the challenge of appropriately timing research on evaluation influence that has the best chance of catching the relationship between mechanisms.

As detailed in the review, a number of other approaches have been adopted in evaluation influence, including surveys and citation analysis. Studies that employed a survey (distinct from case studies that also included a survey) conceptualized influence in terms of survey items reflecting different influence mechanisms (Burr, 2009), or as a set of examples of types of influence events that may occur (Baptiste, 2010). In the case of Christie's (2007) simulated decision-making research, the surveys focused on individual-level influence, and how effective the simulated evaluation data were in changing the existing beliefs of the participants.

Within the evaluation influence literature, there was also a body of research drawing on citation analysis as an approach to assessing the influence of an evaluation. These studies have connected the citation of evaluation documents to evaluation influence, simplifying the understanding of what constitutes influence in order to be able to provide a comprehensive metric. This presents as an innovative approach to studying evaluation influence, albeit one with significant limitations as identified by the researchers themselves (Greenseid & Lawrenz, 2011).

What Does This Research Suggest About the Utility of Evaluation Influence?

Mark (2008) suggests that in order to address core questions of the effects of evaluation practice, researchers should begin with simple research

questions to build a body of evidence. Largely this has not occurred, with much of the literature constituted by conceptually broad case studies across many different areas of study, with much variation in methods, definitions, and procedures. While many researchers have employed influence ideas, this has not occurred in the context of a clear direction for research with implications for evaluation practice. While the use of case studies is not problematic in itself, and indeed influence seems to be a concept well suited to retrospective case studies, the existing body of research on evaluation influence has a number of significant limitations:

- Much of the research is built on the investigation of influence by the same individuals that conducted the evaluation, often based primarily on their experience of undertaking the evaluation (e.g., Diaz-Puente et al., 2008, 2009; Weiss et al., 2005). While these evaluator/researchers are in a position to provide the best knowledge of how an evaluation played out, there is an issue of the potential bias in the way influence is reported. Evaluators should be encouraged to do this work and think about influence, but in the context of a transparent and replicatable research strategy that can reduce the potential for bias;
- The methodological rigor of these studies varies greatly. Some provided only limited description of method, and no clear operational definitions of influence or change mechanisms (Henry & Mark, 2003a). Although some studies reported using case study protocols, there is a need for clear, explicit, and replicatable reporting of method;
- Many studies rely primarily on self-report by organizational stakeholders who may have a direct interest in presenting a narrative of an organization that is receptive to evaluation evidence. While almost all the studies cited used organizational documents along with interviews, there is a need to address directly the desirability of being an evidence-based organization in studying influence.

These issues are similar to other critiques of research on evaluation (Brandon & Fukunaga, 2014; Henry & Mark, 2003b).

While many of the studies discussed in this review provide interesting insights into evaluation influence playing out in specific cases, these studies have limited value in the development of a coherent body of literature to inform evaluation practice. Overall, relatively few studies have findings that directly contribute to the development of evaluation influence, which

is troubling given the prominence of evaluation influence in the evaluation research literature in the context of limited empirical research on evaluation generally (Henry & Mark, 2003b). Some of the studies reviewed set out to examine the relationship between the influences they observed, those that did reported varying degrees of success in following the interaction and interrelationship between different mechanisms. While finding some examples of clear threads of influence, Cheng (2006) and Weiss et al. (2005) reported the exercise challenging due to the intertangling of the threads of influence. Five studies included in the review were more successful in tracing influence, with each beginning their research fairly soon after the evaluation began (Diaz-Puente et al., 2008, 2009; Fjellström, 2007; Gildemyn, 2014; Lehtonen, 2010; Oliver, 2008). This may suggest that researchers should investigate influence alongside or closely following an evaluation in order to capture the individual level mechanisms that tend to begin longer chains of influence.

Evaluation influence has been an important development in the past decade of research on the impact of evaluation. Building on long-standing dissatisfaction with the definitions of use in theory and the research literature, the proponents of this change have made a case for understanding the effects evaluation can have in the broadest sense, in order to enable evaluation researchers to better describe and understand what occurs during and following an evaluation. Evaluation influence represents a new and developing approach to understanding the impact an evaluation has.

Note

1. The author would like to acknowledge and thank the reviewers of this paper for suggesting this categorisation.

References

Alexander, H. (2003). Health-service evaluations: Should we expect the results to change practice? *Evaluation*, 9, 405–414.

Alkin, M. C., Daillak, R., & White, B. (1979). *Using evaluations*. Beverly Hills, CA: Sage.

Alkin, M. C., & Taut, S. M. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, 29, 1–12.

Appleton-Dyer, S., Clinton, J., Carswell, P., & McNeill, R. (2012). Understanding evaluation influence within public sector partnerships: A conceptual model. *American Journal of Evaluation*, 33, 532–546.

Baptiste, L. J. C. (2010). *Process use across evaluation approaches: An application of Q methodology in program evaluation* (Doctor of Philosophy). Kent State University, Kent, OH.

Becker, H. A., Kirkhart, K. E., & Doss, D. (1982). Evaluating evaluation reports. *CEDR Quarterly*, 15, 5–8.

Benjamin, L. M., & Misra, K. (2006). Doing good work: Implications of performance accountability for practice in the nonprofit sector. *International Journal of Rural Management*, 2, 147–162.

Birkeland, S., Murphy-Graham, E., & Weiss, C. H. (2005). Good reasons for ignoring good evaluation: The case of the drug abuse resistance education (D.A.R.E.) program. *Evaluation and Program Planning*, 28, 247–256.

Brandon, P. R., & Fukunaga, L. L. (2014). The state of the empirical research literature on stakeholder involvement in program evaluation. *American Journal of Evaluation*, 35, 26–44.

Burr, E. M. (2009). *Evaluation use and influence among project directors of state GEAR UP grants* (Doctory of Philosophy). University of Tennessee, Knoxville, TN.

Caplan, N. (1976). Social research and national policy: What gets used by whom, for what purposes, and with what effects? *International Social Science Journal*, 28, 187–194.

Caplan, N. (1980). What do we know about knowledge utilization. *New Directions for Program Evaluation*, 5, 1–11.

Caplan, N., Morrison, A., & Stambaugh, R. (1975). *The use of social science knowledge in policy decisions at the national level: A report to respondents*. Ann Arbor, MI: Instititute for Social Research, The University of Michigan.

Caracelli, V. J. (2000). Evaluation use at the threshold of the twenty-first century. *New Directions for Evaluation*, 88, 99–111.

Cheng, S. (2006). *A case study of evaluation use and influence in school settings* (Doctor of Philosophy). University of Minnesota, Minnesota, MN.

Christie, C. A. (2007). Reported influence of evaluation data on decision makers' actions. *American Journal of Evaluation*, 28, 8–25.

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

Cousins, J. B. (2004). Commentary: Minimising evaluation misuse as principled practice. *American Journal of Evaluation*, 25, 391–397.

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

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

Cowley, K., & Good, K. (2010). Using evaluation to refine and focus technical assistance. In S. Harsh, K. Bradley, K. Good, & J. Ross (Eds.), *Capacity building technical assistance: Change agent analyses*. Charleston, SC: Edvantia.

Dahler-Larsen, P. (2001). From programme theory to constructivism: On tragic, magic and competing programmes. *Evaluation*, 7, 331–349.

Diaz-Puente, J. M., Montero, A. C., & Carmenado, I. R. (2009). Empowering communities through evaluation: Some lessons from rural Spain. *Community Development Journal*, 44, 53–67.

Diaz-Puente, J. M., Yague, J. L., & Afonso, A. (2008). Building evaluation capacity in Spain: A case study of rural development and empowerment in the European Union. *Evaluation Review*, 32, 478–506.

Fjellström, M. (2007). The influence of evaluation in higher education: Curriculum development through a deliberative responsive dialogue with stakeholders. *Evaluation Journal of Australasia*, 7, 25–30.

Fleming, M. A. (2011). Attitudes, persuasion, and social influence: Applying social psychology to increase evaluation use. In M. Mark, S. I. Donaldson, & B. Campbell (Eds.), *Social psychology and evaluation* (pp. 212–243). New York, NY: The Guildford Press.

Florio, D. H., Behrmann, M. M., & Goltz, D. L. (1979). What do policy makers think of educational research and evaluation? Or do they? *Educational Evaluation and Policy Analysis*, 1, 61–87.

Frey, K., & Widmer, T. (2011). Revising Swiss policies: The influence of efficiency analyses. *American Journal of Evaluation*, 33, 494–517.

Garcia, J. M. (2008). *A realistic account of evidence-informed tobacco control practice in Ontario public health agencies* (Doctor of philosophy). University of Waterloo, Ontario. Retrieved from http://libdspace.uwaterloo.ca/handle/10012/3638

Gildemyn, M. (2014). Understanding the Influence of Independent Civil Society Monitoring and Evaluation at the District Level: A Case Study of Ghana. *American Journal of Evaluation*. doi: 10.1177/1098214014525257

Greenseid, L. O. (2008). *Using citation analysis methods to assess the influence of STEM education evaluation* (Doctor of Philosophy). University of Minnesota, Minnesota, MN.

Greenseid, L. O., Johnson, K., & Lawrenz, F. (2008). *Beyond evaluation use: Determining the effect of project participation on the influence of NSF program evaluations A citation analysis of the influence of the ATE, CETP, LSC, and MSP-RETA evaluations on the STEM education and evaluation fields*. Retrieved from http://www.cehd.umn.edu/projects/beu/documents/Beyond%20Evaluation%20Use%20-%20Citation%20Analysis%20Report.doc

Greenseid, L. O., & Lawrenz, F. (2011). Using citation anlysis methods to assess the influence of science technology, engineering, and mathematics education evaluations. *American Journal of Evaluation*, *32*, 392–407.

Grob, G. F. (2003). A truely useful bat is one found in the hands of a slugger. *American Journal of Evaluation*, *24*, 499–505.

Heldt, V. W., Braskamp, L. A., & Filbeck, R. (1973). Effects of 'nonrational' influences on decision making in higher education: A simulation test. *Research in Higher Education*, *1*, 163–172.

Henry, G. T. (2000). Why not use? In V. J. Caracelli & H. Preskill (Eds.), *The expanding scope of evaluation use. New Directions for Evaluation* (Vol. 88., pp. 85–98). San Francisco, CA: Jossey-Bass.

Henry, G. T. (2003). Influential evaluations. *American Journal of Evaluation*, *24*, 515–524.

Henry, G. T., & Mark, M. M. (2003a). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, *24*, 293–314.

Henry, G. T., & Mark, M. M. (2003b). Toward an agenda for research on evaluation. *New Directions for Evaluation*, *97*, 69–80.

Hofstetter, C. H., & Alkin, M. C. (2003). Evaluation use revisited. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 197–222). Dordrecht, The Netherlands: Kluwer Acadmic Publishers.

Johnson, R. B. (1998). Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*, *21*, 93–110.

Kirkhart, K. E. (1995). *Consequential validity and an integrated theory of use*. Paper presented at the Evaluation 1995 International Evaluation Conference, Vancouver, BC, November 1995.

Kirkhart, K. E. (2000). Reconceptualising evaluation use: An integrated theory of influence. *New Directions for Evaluation*, *88*, 5–23.

Lawrenz, F., Gullickson, A., & Toal, S. A. (2007). Dissemination: Handmaiden to evaluation use. *American Journal of Evaluation*, *28*, 275–289.

Lawrenz, F., King, J. A., & Ooms, A. (2011). The role of involvement and use in multisite evaluations. *New Directions for Evaluation*, *129*, 49–57.

Ledermann, S. (2012). Exploring the necessary conditions for evaluation use in program change. *American Journal of Evaluation*, *33*, 159–178.

Lehtonen, M. (2010). Indicators as an appraisal technology: Framework for analysing the policy influence of the UK energy sector indicators. Retrieved from http://www.point-eufp7.info/storage/Lehtonen%20-%20EPOS%20article%20 17-09-10.pdf

Leviton, L. C. (2003). Evaluation use: Advances, challenges and applications. *American Journal of Evaluation*, *24*, 525–535.

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

Mark, M. M. (2001). Evaluation's future: Furor, futile, or fertile? *American Journal of Evaluation*, *22*, 457–479.

Mark, M. M. (2003). Toward a comprehensive view of the theory and practice or program and policy evaluation. In S. I. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium*. (pp. 183–204). Hillsdale, NJ: Lawrence Erlbaum.

Mark, M. M. (2008). Building a better evidence-base for evaluation theory. In P. R. Brandon & N. L. Smith (Eds.), *Fundamental issues in evaluation* (pp. 111–134). New York, NY: Guilford.

Mark, M. M. (2011). Toward better research on—And thinking about—Evaluation influence, especially in multistie evaluations. *New Directions for Evaluation*, *129*, 107–119.

Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, *10*, 35–57.

Mark, M. M., & Henry, G. T. (2012). Multiple routes: Evaluation, assisted sensemaking and pathways to betterment. In M. C. Alkin (Ed.), *Evaluation roots: A wider perspectiveof theorists' views and influences* (pp. 144–156). London, England: Sage.

McEathron, M. A. (2008). *Independent science review in natural resource management: Evaluation's role in knowledge use* (Doctor of Philosophy). University of Minnesota, Minnesota, MN.

Morabito, S. M. (2002). Evaluator roles and strategies for expanding evaluation process influence. *American Journal of Evaluation*, *23*, 321–330.

Murphy, D. M. (2007). *Beyond accountability: An empirical study of the factors associated with the use of evaluation for organisational learning in North Carolina's non-profit sector* (Doctor of Philosophy). North Carolina State University, Raleigh, NC.

Oliver, M. L. (2008). *Evaluation of emergency response: Humanitarian aid agencies and evaluation influence* (Doctor of Philosophy in Public Policy). Georgia State University, Atlanta, GA.

Patton, M. Q. (1986). *Utilization-focused evaluation* (2nd ed.). Thousand Oaks, CA: Sage.

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

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (1998). Discovering process use. *Evaluation*, *4*, 225–233.

Patton, M. Q. (2000). Overview: Language matters. In R. K. Hopson (Ed.), *How and why language matters in evaluation, new directions in evaluation* (Vol. 86). San Francisco, CA: Jossey-Bass.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.

Patton, M. Q., Grimes, P. S., Guthrie, K., 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, NY: D.C. Health.

Poth, C. N. (2008). *Promoting evaluation use within dynamic organisations: A case study examining evaluator behaviour* (Doctor of Philosophy). Queen's University, Ontario, Canada.

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

Rebolloso, E., Baltasar, F., & Canton, P. (2005). The influence of evaluation on changing management systems in educational institutions. *Evaluation*, *11*, 463–479.

Rich, R. F. (1977). Uses of social science information by federal bureaucrats: Knowledge for action versus knowledge for understanding. In C. H. Weiss (Ed.), *Using social research in public policy making*. Lexington, KY: Lexington Books.

Roseland, D., Greenseid, L. O., Volkov, B. B., & Lawrenz, F. (2011). Documenting the impact of multisite evaluations on the science, technology, engineering, and mathematics field. *New Directions for Evaluation*, *129*, 39–48.

Sandison, P. (2006). The utilisation of evaluations. *Active Learning Network for Accountability and Performance in Humanitarian Action Review of Humanitarian Action in 2005*. Retrieved from http://www.alnap.org/resource/5225.aspx

Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, *18*, 195–209.

Valovirta, V. (2002). Evaluation utilization as argumentation. *Evaluation*, *8*, 60–80.

Vataja, K. (2011). Developing work and services in public sector organizations through evaluation. *Scandinavian Journal of Public Administration*, *15*, 47–67.

Weets, K. (2008). *How effective are performance audits? A multiple case study within the local audit office of Rotterdam*. Paper presented at the 5th International Conference on Accounting, Auditing & Management in Public Sector Reforms, Amsterdam, 3-5 September.

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

Weiss, C. H. (1987). Where politics and evaluation research meet. In D. Plaumbo (Ed.), *The politics of program evaluation*. Newbury Park, CA: Sage.

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

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

Weiss, C. H., & Bucuvalas, M. J. (1977). The challenge of social research to decision-making. In C. H. Weiss (Ed.), *Using social science research in public policy making*. Lexington, KY: Lexington Books.

Weiss, C. H., & Bucuvalas, M. J. (1980). Truth tests and utility tests: Decision-makers' frames of reference for social science research. *American Sociological Review*, *45*, 302–313.

Weiss, C. H., Murphy-Graham, E., & Birkeland, S. (2005). An alternative route to policy influence: How evaluations affect D.A.R.E. *American Journal of Evaluation*, *26*, 12–30.

Author Biography

James Leslie Herbert is a Post-Doctoral Research Fellow at the Australian Centre for Child Protection at the University of South Australia. He completed his PhD on practitioner oriented evaluation at the School of Social Sciences and Psychology at the University of Western Sydney.