The Historical Development of Evaluation Use

$\$$SAGE

Marvin C. Alkin[1] and Jean A. King[2]

Abstract

The first article in this series traces the initial development of the concept of evaluation use. As a field, evaluation has always paid attention to the potential for use, both in decision-making and in changing people's thinking. The broad history of the field as we know it today stemmed from two streams: one focused on tests and measurement, primarily in education, and a second focused on social research methods, primarily concerning knowledge utilization. Evaluation use had its roots in both streams, resulting in three broad categories for discussing the use of evaluation findings: instrumental use, conceptual use or enlightenment, and symbolic use. The additional category of process use, added years later, highlighted the potential utility of people's participation in the evaluation process.

Preface

This article is the first of three tracing the historical development of the concept of evaluation utilization/use.[1] In the first article of the series, we highlight the centrality of utility since the field's beginning, documenting two developmental streams, one stemming from educational testing and measurement and the other from the social sciences. We also describe evolving thought on evaluation use, competing perspectives about its prevalence, and its traditional categories in the literature, including the use of evaluation findings and its process.

The series' second article will address three purposes: explicate definitions of evaluation use and its undesirable companion, misuse; discuss a broadened concept of evaluation impact that includes both use and evaluation influence; and finally, identify the factors that research over time has shown to be associated with evaluation use.

The final article in the series will present the evolution of theories of and research on evaluation use. First, we review the prescriptive theories presented on the use branch of Alkin's evaluation

[1] University of California, Los Angeles, Los Angeles, CA, USA
[2] University of Minnesota, Minneapolis, MN, USA

Corresponding Author:
Marvin C. Alkin, University of California, Los Angeles, Box 951521, Los Angeles, CA 90095, USA.
Email: alkin@gseis.ucla.edu

theory tree (Alkin, 2013). Then, we summarize and critique the empirical research on evaluation use and discuss the influence of competing understandings of use.

Origins and Conceptual Grounding of Use

The history of evaluation use is necessarily part of the general history of program evaluation, and to understand its development and meaning, it is helpful to ground it within a historical context. Shadish and Luellen (2005) have commented that the history of evaluation is as old as the history of human activity: "Humans (a) identify a problem, (b) generate and implement alternatives to reduce its symptoms, (c) evaluate those alternatives, and then (d) adopt those that results suggest will reduce the problem satisfactorily" (p. 183). Namely, people engage in activities they refer to as evaluation in order to foster "use."

Shadish and Luellen provide evidence of evaluation dating back thousands of years, including the discussion of the evaluation of the Hebrew diet regimen in Chapter 1 of the Biblical Book of Daniel (see also Patton, 2008) and personnel evaluation in China more than 4,000 years ago. These two examples provide the basis for understanding evaluation's deep historical roots. Many scholars affirm that evaluation use has been a matter of concern almost as long as people have practiced evaluation itself, including Riecken (1953/1972), who discusses this in depth. The explanation is quite simple. Evaluation's raison d'être is the contribution it makes to better program operations. It is intended to be a practical craft. Weiss (1972) made the point well:

> The basic rationale for evaluation is that it provides information for action. Its primary justification is that it contributes to the rationalization of decision-making . . . [U]nless it gains serious hearing when program decisions are made, it fails in its major purpose. (p. 318)

Cronbach and Suppes (1969) provided further basis for understanding that the expectation of use is an important part of evaluation. They referred to both evaluation and research as forms of "disciplined inquiry," noting that there were differences between these two types of disciplined inquiry. They distinguished "decision-oriented" from "conclusion-oriented" investigations. A decision-oriented study is one in which "the investigator is asked to provide information wanted by a decision maker: a school administrator, a governmental policy-maker, the manager of a project to develop a new biology textbook, or the like" (1969, p. 20). The conclusion-oriented study, by contrast, takes its direction from the investigator's commitments and hunches and is guided by the desire to add to a disciplinary base. Thus, evaluations address the here and the now (i.e., *this* program at *this* time) and attempt to provide insights intended to lead potentially to the use of findings for program improvement either in the form of decisions or better program understanding.

Michael Scriven's (1967) definition of evaluation—the "judgment of merit and worth" provided further insight. Merit refers to intrinsic value; worth refers to extrinsic value or goodness. Some things may be intrinsically good but not of value to an individual or to an entity. More specifically, a program may have merit (i.e., it is considered to be of good quality), but it may not have value for a particular context. In essence, Scriven told us that evaluation is context driven. If the study is in a different place, it is a different context. And even if it is the same place, but at another time, the context may have changed. The notion of "worth" is, in essence, the idea that evaluations must have "utility"—the potential to be used in the given context at a given time.

Two Streams of Evaluation History as Related to Use

The broad history of evaluation as we know it today follows two streams. There is one stream, most heavily identified with the education field, that focuses on tests and measurements. There is a second stream, a social science stream that focuses on the use of social research methodology. The examples

provided above by Shadish and Luellen illustrate these two streams. The tests administered in the China personnel evaluation clearly resulted in making appointments—that was their purpose. The experiment described in the Book of Daniel led to the use of the evaluation information and dietary changes.

These two streams were quite distinct in the early history of evaluation use but in modern times have converged to a great extent. The merger of The Evaluation Research Society and The Evaluation Network to form the American Evaluation Association in 1986 may have marked a turning point. We will examine the early history of each of these streams and then turn to an additional concept that reflects a more recent development.

Measurement Stream

Travers (1983) provided ample documentation of evaluation activity in the testing stream from the mid-1800s to the early 1900s. He noted Horace Mann's early work in preparing annual reports over 12 years to the Board of Education of the Commonwealth of Massachusetts. These reports focused on a number of aspects of schooling. They, in turn, stimulated the development of The Boston Survey, in which a sample of Boston students were tested on a variety of school subject areas. Travers (1983) noted that the testing was discontinued after 2 years because no use was made of the test results—an early instance of recorded "nonuse." In 1897, Joseph Rice reported having carried out an assessment program using standardized tests in a number of U.S. school systems. He even proposed an early version of what we today consider to be adversary or judicial evaluation (Fitzpatrick, Sanders, & Worthen, 2004).

In the early 1900s, the writings of Edward Lee Thorndike provided a particular stimulus to the school testing movement. Thorndike wrote *An Introduction to the Theory of Mental and Social Measurements* (1913), subsequent books on other measurement topics, and numerous journal articles on school testing generally and on testing for the specific subject areas. Stimulated by these technical achievements, over 40 large school districts had school research units conducting large-scale assessments of student achievement by World War I (Madaus, Arasian, & Kellaghan, 1980). These assessments served a variety of purposes, including assessing school performance. Many, if not most, large districts still have such units. As Fitzpatrick, Sanders, and Worthen (2004) noted,

> During this period, measurement and evaluation were regarded as nearly synonymous, and the term *evaluation* was most often used to mean the assigning of grades or summarizing of student performance on tests. The concept of evaluation, as we know it today, was still evolving. (p. 14)

The 1930s was a time of another influential event for the measurement stream of evaluation. John Dewey's writings among others had provided the impetus for what was known as the progressive education movement. Schools experimented with new pedagogy and curricula, but doubters criticized their worth as compared to the traditional Carnegie unit curricula. This led to the well-known Eight Year Study at the University of Chicago (Smith & Tyler, 1942). Systematic objectives-based (or criterion-referenced) testing procedures were developed to evaluate these curricula. The Eight Year Study presented a use emphasis, as the authors acknowledged the need to "provide for the training of teaching and school officers in the interpretation and use of evaluation results" (Madaus & Stufflebeam, 1989, p. 113). These testing procedures have had continuing professional influence. Witness their presence even today in, for example, the National Assessment of Educational Progress, first headed by Ralph Tyler, one of the leaders of the Eight Year Study.

Two decades later, Cronbach's article, "Evaluation for Course Improvement" (1963), further enhanced interest in evaluation use. Cronbach considered the variety of new mathematics and science curriculum projects in response to the Union of Soviet Socialist Republics's launching of

Sputnik 1. In this influential article, he faulted the adequacy of the evaluations of these projects. He identified a course improvement objective for evaluation. By this, he meant that the appropriate role of evaluation was not only to make final judgments about the efficacy of the curricula but also to provide information that would assist in making modifications of the courses under development. Cronbach's idea that course improvement was an appropriate outcome of evaluation activity became the basis for the formative/summative distinction that Scriven (1967) subsequently made. In essence, Cronbach's prescriptive formulation extended the notion of appropriate evaluation use beyond simply looking at final program outcomes. Owing to Cronbach's esteem as a methodologist, the impact and stimulus of this article were substantial.

The context that subsequently focused greater attention on evaluation and the apparent lack of success led to increased concern about use. The 1960s was a period evidenced by a great increase in the number of evaluations. Major social programs were initiated that often mandated end-of-year evaluation reports. The mid-1960s were years when the U.S. Congress enacted numerous "Great Society" programs. In education, several sections of the 1965 Elementary and Secondary Education Act provided additional federal resources for serving the programmatic needs of educationally disadvantaged youth. The Act required that school district recipients of funding engage in, or contract for, evaluations of these programs to demonstrate the impact of the federal funds. As a part of that act, Senator Robert Kennedy and others embraced the idea that an important purpose of evaluations was its use by parents and local educators to improve school programs. Thus, a requirement in the Act mandated the establishment of local School Site Councils consisting of parents, teachers, and administrators. Moreover, evaluators were required to submit a report to these Councils. The notion of evaluation as relevant and potentially useful to the various people who have a stake in the program, thus, was given political impetus.

If the good news was that the attention to evaluation in these Great Society programs attracted many measurement experts or academic researchers, the bad news was that these researchers often failed to understand properly the contexts in which evaluations were to take place. Their attempts at simply implementing testing procedures or attempting to conduct research in schools by and large failed. The result was a multitude of school evaluation reports that lacked perceived relevance to programs or meaning to program personnel. This was particularly true for evaluations conducted at the local program level. Thus, many evaluation reports sat on the shelf, fulfilling the requirement of evaluation, but generally ignored or disregarded by those with the authority to do something with their results. Testing might have been conducted, but there was a lack of understanding of the differences between testing and evaluation.

This lack of understanding was evidenced and further reflected in subsequent changes made in measurement textbooks. A Google Ngram (Michel et al., 2010) created by searching for the words *measurement* and *evaluation* in books from 1940 to 2000 documents that the percentage of published books using the word *evaluation* doubled between 1965 and 1980, while the percentage of books including the word *measurement* remained roughly the same.[2] Indeed, several leaders working in the field at that time seemed to confirm our view that a number of measurement texts simply added the word *evaluation* to their titles with little, if any, change in content (David Berliner, personal communication, April 21, 2016; James Popham, personal communication, April 25, 2016; John McNeil, personal communication, May 2, 2016).

This general lack of confidence in existing methods for conducting evaluation led scholars such as Scriven and Stake to suggest better ways to conduct program evaluation. Also, a number of authors, including Stufflebeam (1971), Provus (1971), Cooley and Bickel (1986), Worthen and Sanders (1987), among others, produced textbooks prescribing what were presumed to be more relevant ways of conducting evaluation, that is, "prescriptive theories" of evaluation. Patton, whose work was really a fusion of the two streams, also published the first edition of his book on utilization-focused evaluation (1978).

Social Science Stream

The social science stream of program evaluation history was primarily defined by the application of social science research methods. Shadish and Luellen (2005) have noted that substantial growth and refinement of empirical research methods contributed to increased instances of evaluation activity. Caro (1971) cited studies conducted by Elton Mayo, Fritz Roethlisberger, William Dixon, and Stuart Dodd in the years 1920–1950 as examples (pp. 4–5).

Yet these studies in the Franklin Roosevelt era focused on using social science research to produce what is referred to as "knowledge use," an apt term since this literature is concerned with the creation of conclusion-oriented information, that is, knowledge. Such knowledge is generally the product of research conducted at a given time. Alternatively, it may refer to the use of research information generally collected (or acquired) at another time and potentially applicable for shedding light on a current situation. Generally, social scientists hoped that their efforts would influence the policy development process and ultimately contribute to the improvement of societal functioning and human welfare. Typically, in many of these studies, there was no direct focus on determining the quality of a program. As Fitzpatrick et al. (2004) note, most of these social scientists "... pursued applied research at the intersection of the agency's needs and their personal interests, thus, sociologists pursued questions of interest to the discipline of sociology and the agency" (p. 33).

As noted previously, the 1960s and early 1970s introduced another era of serious attention to evaluation. As was true in the measurement branch of evaluation, the advent of Great Society programs affected the social research branch. The abundance of federal programs (e.g., The Manpower Development and Training Act, The Economic Opportunity Act) overwhelmed the existing capacity to perform these evaluations. Social research academics tried to fill the void. As in the prior period, much research was highly discipline based and not perceived as relevant by stakeholders.

Thus, despite increased federal funding for social research programs nationwide, in the 1960s and 1970s, social scientists and government officials remained concerned that research efforts went largely unnoticed and that policies were often debated and passed in spite of contrary research findings (Bogenschneider & Corbitt, 2010; Gray & Lowery, 2000; Knorr, 1980). The perception that there was little indication that knowledge had impact on policy decision makers, that is, this awareness of potential nonuse, provided the impetus for further research (Weiss, 1989).

Initial systematic consideration of the concept of knowledge utilization was conducted by Nathan Caplan at the Center for Research on Utilization of Scientific Knowledge at the University of Michigan and by Robert Rich at the National Opinion Research Center. Caplan, Morrison, and Stambaugh (1975) and their associates examined the ways in which social science knowledge influenced federal decision-making processes. In their work, they made an early distinction between "hard" and "soft" knowledge. The former was identified as research based, usually quantitative, and couched in scientific language, and the latter as nonresearch-based, qualitative, and couched in lay language. They noted that "our data suggest that there is widespread use of soft information and that its impact on policy, although often indirect, may be great or even greater than the impact of hard information" (p. 47).

At about the same time, Rich, also looking at the use of data in federal decision-making, found it helpful to distinguish knowledge for action (which he referred to as "instrumental utilization") and knowledge for understanding (conceptual utilization; Rich, 1977). The former, which we now refer to as instrumental use, indicated instances where respondents in the study could document the specific way they had used the social science knowledge for decision-making or problem-solving purposes. Conceptual use indicated instances in which knowledge about an issue influenced a policy maker's thinking, although the information was not put to a specific, documentable action. Pelz

(1978) synthesized the research of Caplan and Rich, highlighting the conceptual difficulties in measuring knowledge use.

Carol Weiss (1977) amplified understanding of conceptual use by arguing that rather than playing a deliberate, measurable role in decision-making, social science knowledge "enlightens" policy makers. According to Weiss, policy makers valued studies that prompted them to look at issues differently, justified their ideas for reforms, challenged the status quo, and suggested the need for change. Thus, the term *enlightenment* became part of the standard verbiage in the knowledge literature and subsequently in the evaluation literature more broadly. The notion of enlightenment suggested that knowledge from social science research more frequently was not the sole source leading to policy decisions—rather, it "informed" policy decisions. Moreover, the knowledge might be employed in more subtle ways, what Weiss (1977) referred to as "interactive." In this model, knowledge is used in conjunction with a decision maker's personal insights, experience, and information communicated.

In addition to conceptual and instrumental use, Karin Knorr (1977) identified what she called a third mode of use of social science data referred to as "symbolic use." She considered two forms of symbolic use. One instance is where a government official signals by announcing an evaluation that something is being done about a problem, while proper actions that should have been taken are postponed or ignored altogether. A second, more common aspect of symbolic use that she identified was a legimating role. In these situations, data are used to publicly support a decision made on a different basis or on an already held opinion.

Pelz noted that the distinction among the three use categories was not sharp. The examples he provided were compelling:

> If the evidence *persuades* a decision maker to adopt option A rather than B, the use is clearly instrumental. If he has already adopted option B, however, and the evidence reinforces his belief in this option, the use is more properly called conceptual . . . The demarcation between conceptual and symbolic is also blurred. If information serves to confirm the decision maker's *own judgment* of the situation, we have a conceptual use. If the evidence helps him to justify his position to *someone else*, such as legislative committee or a public group, the use is symbolic. (1978, p. 352)

Another important contribution from those engaged in the study of knowledge use was the linkage of social science knowledge with "ordinary knowledge." Such knowledge emerges from everyday personal experiences, rather than from social science evidence, to inform the decision-making process and is sought in "common sense, casual empiricism, or thoughtful speculation and analysis" (Lindblom & Cohen, 1979, p. 12). "What differentiates ordinary knowledge from social science knowledge is how it is verified by the user. Thus, it can be incorrect or false, as it derives from nonsystematic observation and verification strategies" (Lindblom & Cohen, 1979, p. 201). In many respects, this is a similar idea to that which Weiss proposed in her discussion of the interactive use of knowledge and that Caplan et al. (1975) had suggested earlier.

Kennedy (1983) extended the concept of ordinary knowledge to the work environment by identifying "working knowledge" as "context dependent, functioning as a filter to interpret incoming, new social science evidence and to judge the validity and usefulness of the new evidence" (p. 202)

This research on knowledge utilization is highly relevant to an understanding of the development of the concept of evaluation use. Many of the burgeoning ideas related to knowledge utilization had great influence on and served as starting points for research on evaluation use by scholars in both streams. However, despite the influence of this earlier work, Leviton and Hughes (1981) cautioned against drawing too great an inference from the knowledge utilization literature, as program evaluation differed from social research commonly used in government because governments were more typically bound by political constraints.

The Evolving Body of Thought on Evaluation Use

From the early 1970s forward, the two evaluation streams appear to converge—at least in relation to the development, study, and research on evaluation use. The early work of Carol Weiss, in particular her chapter "Utilization of Evaluation: Toward Comparative Study" (Weiss, 1972), heavily influenced the concern for more directed attention to studying and improving evaluation use. This article was a field-changing classic that set the tone for a research agenda on evaluation use for many years.

Competing Perceptions About the Prevalence of Use

In her call for a focus on evaluation utilization issues (1972), Carol Weiss cited the need for research because of what she perceived as the prevalence of nonuse of evaluation information. She indicated that while some instances of effective utilization existed, the rate of nonutilization was much greater. This issue also had been considered by Davis and Salasin (1975), who opined that use occurred more frequently than noted in the literature but failed to get recognized because expectations were set too high and time frames too short. But the issue remained unresolved and indeed persisted. Guba published an article entitled "The Failure of Educational Evaluation" (1969), and Rippey (1973) commented:

> At the moment, there seems to be no evidence that evaluation, although the law of the land, contributes anything to educational practice other than headaches for the researcher, threats for the innovators, and depressing articles for journals devoted to evaluation. (p. 9)

The issue of the prevalence of use was addressed at a small, 3-day meeting at Malibu, CA, in 1987 (Alkin, 1990). In particular, Michael Quinn Patton and Carol Weiss engaged in heated discussion regarding the extent to which evaluation use occurred. The disagreement reerupted in a more vigorous fashion in major presentations that each made at the 1987 Annual Meeting of the American Evaluation Association (Patton, 1988; Weiss, 1988). While both agreed that there were a number of instances where evaluations led to decisions (including program modifications and changes in attitudes), both seemed surprised that, in light of the Malibu discussions, they did not better understand each other's views. Weiss viewed as an indicator of prevalence of use whether the evaluation was the *only* basis for the decision or even the overriding concern. Patton employed a more generous definition of influence (not as that term is used today in the literature).

   Alkin (1990) commented on this disagreement and noted that the main issue defining their differences was predominantly related to the scope of evaluations that each addressed. Weiss had conducted evaluations that primarily focused on large agencies, legislatures, the Congress, and so on. These evaluations had high visibility and multiple constituencies. Thus, they tended to a far greater extent to "accrete," the term that Weiss uses to refer to the gradual aggregation of information (Weiss, 1980). Patton, on the other hand, had tended to engage in smaller, local program-oriented evaluations, which were generally less politically "loaded," less visible, and had fewer people involved in making decisions. Discussions in these instances might tend to be less "go/no go" and more formative in nature—small program modifications. Differences in the perception of extent of evaluation use might very well be partially related to the program size. And a key element was the extent to which the context of different program types (e.g., governmental level, size, political) affected the degree to which people perceived the evaluation as being used. It is important to note that given subsequent research on evaluation use and the extent to which the concept has become visible, there is a greater prevalence of use now than when that debate took place.

Process Use

Another important development in the history of evaluation use was the addition of process use. The evaluation literature had referred to the idea of process use for some time (e.g., Rippey, 1973). King and Pechman's (1984) case study of a school district's research and evaluation unit used the term *use process* to describe how "an awareness of this process on both evaluators' and users' parts may enable the evaluation community ... to look upon the completion of an evaluation as a worthwhile activity ... " (p. 251). Recognizing that the evaluation process itself could have impact, Cousins and Earl (1992), Greene (1988), and King (1988) also discussed the notion of *process use* implicitly. However, it was Michael Patton who named the concept and brought the idea of process use to the attention of the evaluation community (1994). In a 1998 article in the *American Journal of Evaluation*, he related how the idea arose and demanded his attention:

> I have asked intended users about actual use. What I would typically hear was something like: "Yes, the findings were helpful in this way and that, and here's what we did with them." If there had been recommendations, I would ask what subsequent actions, if any, followed. But, beyond the focus of findings and recommendations, what they almost inevitably added was something to the effect that "it wasn't really the findings that were so important in the end, it was going through the process." (Patton, 1998, p. 225)

Thus, *process use* emerged as a term to address the ways that activities through which an evaluation was conducted—rather than its findings—affected individuals and the organization. Patton formally defined process use as "individual changes in thinking and behavior ... that occur among those involved in the evaluation as a result of the learning that occurs during the evaluation process" (1997, p. 90). He further elaborated that "changes in program or organizational procedures and culture may also be manifestations of process impacts" (Patton, 1998, p. 225). If, for example, the program was altered due to the thinking process inspired by engaging in the evaluation, then this was an indication of process use. Or the evaluator might have indicated that certain kinds of data would be collected as part of the evaluation. An understanding of the measures to be used might lead to a recognition of project deficiencies before any evaluation data had been collected.

Some writers have mistakenly identified process use as a kind of use parallel to instrumental or conceptual use. It is important to note that process use is not an additional use category. Rather, it refers to the source of the stimulus—that is, the process and not the findings. Thus, just as with findings use, process use has both instrumental and conceptual consequences (Alkin & Taut, 2002). That is, the process may lead to decisions or program actions taking place or to changes in thinking and understanding about oneself, the program, or the organization. Thus, a distinction may be made between *findings use* and *process use*, each being employed instrumentally or conceptually. Further, the symbolic use and legitimative use described earlier can and do emanate from the evaluation process. Legitimative use, because it depends on the findings of the evaluation to rationalize an earlier decision, is within the findings use category; so-called symbolic use, which depends only on the process being conducted, is in the process use category (see Figure 1).

Furthermore, process use, like findings use, can take place at different times. Instrumental or conceptual process use may occur during the course of the evaluation or at its conclusion. In the latter instance, one may consider the net result and impact of the participation in the complete evaluation process. In a participatory evaluation, for example, one may expect that evaluation skills are acquired by the conclusion of the evaluation process (instrumental process use) or that people's attitudes have changed that can be applied to other situations (conceptual process use).

Other authors (Cousins, 2007; Owen & Lambert, 1995; Preskill & Torres, 1999, 2000) have extended the idea of process use to the impact on the team and organizational levels. Amo and Cousins (2007) commented on the mechanism that transforms process use to organizational capacity building
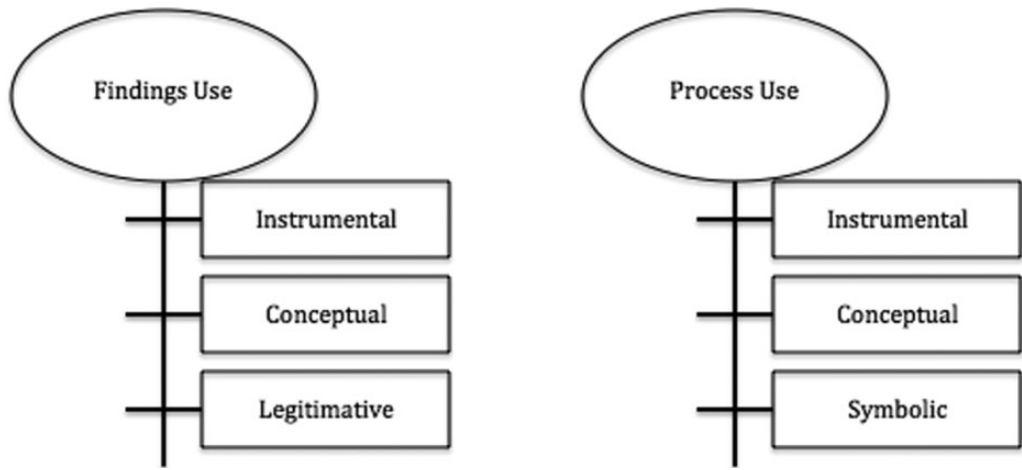
Figure 1. Types of use associated with findings and process. From "Unbundling Evaluation Use," by M. C. Alkin and S. M. Taut, 2002, *Studies in Educational Evaluation*, 29, p. 7. Copyright 2003 by Elsevier Science Ltd. Adapted with permission.

and ultimately to further evaluation use. They wrote, "Process use ... is thought to enhance organizational readiness for evaluation through augmenting organizational capacity to do and use evaluation" (p. 6). The impact of the evaluation process on organization development may occur during the evaluation process or at its conclusion. Likewise, increased appreciation of the value of engaging in the evaluation process may be acquired both during the process as well as at its conclusion.

Typically, evaluators may think of process use as "incidental," that is, an evaluation occurred, and as a result of people's participation, some form of process use occurred. This is exemplified by the earlier quotation in which Michael Patton described how he uncovered process use. As evaluators became more concerned with organizational learning and increasing organizational capacity, the creation of evaluation agendas to *purposefully* attain these goals became more frequent. "Planned process use" (King, 2007) emerged as a potential means of creating a sustainable organizational process by building on people's engagement in evaluation activities. On the developmental timeline, scholarly discussion of the concept of evaluation capacity building emerged a few years following that of process use (Gilliam et al., 2003; McDonald, Rogers, & Kefford, 2003; Stevenson, Florin, Mills, & Andrade, 2002; Stockdill, Baizerman, & Compton, 2002).

Such purposeful process use has an ethical dimension. As Patton (1998) wrote,

> ... ours is not a neutral activity or profession. Precisely because of this potential power of process use— power to influence things for either good or ill—we have to worry about the admonition to first do no harm. (p. 232)

We must recognize that when we seek to foster process use, the goals should reflect the program's values and not our values or those that we evaluators think people should have.

Summary and Conclusion

At its core, the practice of evaluation is concerned with the utility of both its process and its results, making evaluation use an integral and grounding feature of evaluation history. This article has traced the origins of evaluation use from two parallel streams: one from educational measurement and the

perceived need to provide useful information to decision makers, and the other from social science research and the potential of well-crafted knowledge to inform decision-making. Over time, three commonly accepted categories of use based on evaluation findings emerged—instrumental, conceptual, and symbolic. Later an additional category of process use based on what people gain through participation in an evaluation was added to the lexicon; the three categories of findings use applied equally well to process use. Broadening consideration about the use of evaluation to the organizational level expanded the concept, including thoughts on how to create evaluation capacity through the evaluation process.

The second article in this series tracing the historical evolution of evaluation use will begin by examining definitions of evaluation use and misuse and then discuss the origin and effects of the broader concept of evaluation influence. It will also explore the factors that have been empirically supported as associated with evaluation use.

Declaration of Conflicting Interests

Funding

Notes

1. We subsequently will employ the word *use* instead of *utilization*. The literature initially referred to the concept of evaluation utilization, adapting the term from the field of knowledge utilization. In the early 1980s, the evaluation community tended to substitute the word use for utilization. In part, this was a function of the negative loading attached to the word *utilization* reflected in Carol Weiss' feeling about lack of utilization that will be discussed in the article. The term had come to imply the necessity that evaluation be the only prevalent basis for action. Another consideration for embracing the term *use* was the term itself. As Weiss (1979) notes, utilization "embodies an inappropriate imagery ... because of its overtones of instrumental episodic application. People don't utilize research the way that they utilize a hammer" (p. 2).
2. Our sincere thanks to Alana Kinarsky, University of California, Los Angeles, for creating the Ngram.

References

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

Alkin, M. C. (Ed.). (2013). *Evaluation roots: A wider perspective of theorists' views and influences*. Los Angeles, CA: Sage.

Alkin, M. C., & Taut, S. M. (2002). Unbundling evaluation use. *Studies in Educational Evaluation*, *29*, 1–12.

Amo, C., & Cousins, J. B. (2007). Going through the process: An examination of the operationalization of process use in empirical research on evaluation. *New Directions for Evaluation*, *116*, 5–26.

Bogenschneider, K., & Corbitt, T. (2010). *Evidence-based policymaking: Insights from policy-minded researchers and research-minded policymakers*. New York, NY: Routledge.

Caplan, N., Morrison, A., & Stambaugh, R. (1975). *The use of social science research at the national level*. Ann Arbor, MI: Institute for Social Research.

Caro, F. G. (1971). Evaluation research: An overview. In F. G. Caro (Ed.), *Readings in evaluation research* (Vol. 6, pp. 1–34). New York, NY: Russell Sage Foundation.

Cooley, W. W., & Bickel, W. E. (1986). *Decision-oriented educational research* (Vol. 11). Boston, MA: Kluwer-Nijoff.

Cousins, J. B. (Ed.) (2007). Process use in theory, research, and practice. *New Directions in Evaluation*, *116*.

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

Cronbach, L. J. (1963). Evaluation for course improvement. *Teachers College Record*, *64*, 672–683.

Cronbach, L. J., & Suppes, P. (1969). *Research for tomorrow's schools: Disciplined inquiry for education*. New York, NY: MacMillan.

Davis, H. R., & Salasin, S. E. (1975). The utilization of evaluation. *Handbook of Evaluation Research*, *1*, 621–666.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines*. Upper Saddle River, NJ: Pearson Education.

Gilliam, A., Barrington, T., Davis, D., Lacson, R., Uhl, G., & Phoenix, U. (2003). Building evaluation capacity for HIV prevention programs. *Evaluation and Program Planning*, *26*, 133–142.

Gray, V., & Lowery, D. (2000). Where do policy ideas come from? A study of Minnesota legislators and staffers. *Journal of Public Administration, Research and Theory*, *10*, 573–597.

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

Guba, E. G. (1969). The failure of educational evaluation. *Educational Technology*, *9*, 29–38.

Kennedy, M. M. (1983). Working knowledge. *Science Communication*, *5*, 193–211.

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

King, J. A. (2007). Developing evaluation capacity through process use. *New Directions for Evaluation*, *116*, 45–59.

King, J. A., & Pechman, E. M. (1984). Pinning a wave to the shore: Conceptualizing school evaluation use. *Educational Evaluation and Policy Analysis*, *6*, 241–251.

Knorr, K. D. (1977). Policymakers' use of social science knowledge: symbolic or instrumental? In C. Weiss (Ed.), *Using social research in public policy making* (pp. 165–182). Lexington, MA: D. C. Health.

Knorr, K. D. (1980). The gap between knowledge and policy. In S. Nagel (Ed.), *Improving policy analysis* (pp. 219–233). Beverly Hills, CA: Sage.

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

Lindblom, C. E., & Cohen, D. K. (1979). *Usable knowledge: Social science and social problem solving* (vol. *21*). New Haven, CT: Yale University Press.

Madaus, G. F., Airasian, P. W., & Kellaghan, T. (1980). *School effectiveness: A reassessment of the evidence*. New York, NY: McGraw-Hill.

Madaus, G. F., & Stufflebeam, D. L. (1989). *Educational evaluation: Classic works of Ralph W. Tyler*. Boston, MA: Kluwer Academic.

McDonald, B., Rogers, P., & Kefford, B. (2003). Teaching people to fish? Building the evaluation capability of public sector organization. *Evaluation*, *9*, 9–29.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Brockman, W., . . . Aiden, E. L. (2010). Quantitative analysis of culture using millions of digitized books. *Science*, *331*, 176–182.

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

Patton, M. Q. (1978). *Utilization-focused evaluation*. Beverly Hills, CA: Sage.

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

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

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

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

Pelz, D. C. (1978). Some expanded perspectives on use of social science in public policy. In J. M. Yinger & S. J. Cutler (Eds.), *Major social issues: A multidisciplinary view* (pp. 346–357). New York, NY: Free Press.

Preskill, H., & Torres, R. T. (1999). *Evaluation inquiry for learning in organizations*. Thousand Oaks, CA: Sage.

Preskill, H., & Torres, R. T. (2000). The learning dimension of evaluation use. *New Directions for Evaluation*, *88*, 25–37.

Provus, M. (1971). *Discrepancy evaluation*. Berkeley, CA: McCutchan.

Rich, R. F. (1977). Measuring knowledge utilization: Processes and outcomes. *Knowledge and Policy*, *10*, 11–24.

Riecken, H. (1972). Memorandum on program evaluation. In C. H. Weiss (Ed.), *Evaluating action programs: Readings in social action and education* (pp. 85–104). Boston, MA: Allyn & Bacon. (Original work published 1953)

Rippey, R. M. (1973). The nature of transactional evaluation. In R. M. Rippey (Ed.), *Studies in transactional evaluation* (pp. 1–16). Berkeley, CA: McCutchan.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. 1, pp. 39–83). Chicago, IL: Rand McNally.

Shadish, W. R., & Luellen, J. K. (2005). History of evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 183–186). Thousand Oaks, CA: Sage.

Smith, E. R., & Tyler, R. W. (1942). *Appraising and recording student progress. Vol. III. Adventure in American Education*. New York, NY: Harper & Bros.

Stevenson, J. F., Florin, P., Mills, D. S., & Andrade, M. (2002). Building evaluation capacity in human service organizations: A case study. *Evaluation and Program Planning*, *25*, 233–243.

Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, *93*, 7–26.

Stufflebeam, D. I. (1971). The relevance of the CIPP evaluation model for educational accountability. *Journal of Research and Development in Education*, *5*, 19–25.

Thorndike, E. L. (1913). *An introduction to the theory of mental and social measurements*. New York, NY: Teacher's College, Columbia University.

Travers, R. M. (1983). *How research has changed American schools. A history from 1840 to the present*. Kalamazoo, MI: Mythos Press.

Weiss, C. H. (1972). Utilization of evaluation: Toward comparative study. In C. H. Weiss (Ed.), *Evaluating action programs: Readings in social action and education* (pp. 318–326). Boston, MA: Allyn and Bacon.

Weiss, C. H. (Ed.). (1977). *Using social research in public policy making*. Lanham, MD: Lexington Books.

Weiss, C. H. (1979). *Conceptual issues in measuring the utilization of research and evaluation*. Paper presented at the Annual Meeting of the Evaluation Research Society, Minneapolis, MN.

Weiss, C. H. (1980). Knowledge creep and decision accretion. *Knowledge: Creation, Diffusion and Utilization*, *1*, 381–404.

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

Weiss, C. H. (1989). Congressional committees as users of analysis. *Journal of Policy Analysis and Management*, *8*, 411–431.

Worthen, B. R., & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. White Plains, NY: Longman.