Metaevaluation as a Means of Examining Evaluation Influence

Monica LaBelle Oliver
Evaluation Consultant, Atlanta, GA

ABSTRACT: Understanding how (or whether) an organization uses its evaluation findings can help that organization and others like it to evaluate more purposefully and effectively. Understanding the broader influence evaluations have on the organization and in its operating environment holds even greater potential. Metaevaluation is an appropriate approach for determining how effectively evaluations have served the purpose of program improvement. They can potentially expose other intended or unintended ways in which the evaluation influenced the organization and its environs. Using Henry and Mark's (2003) taxonomy of evaluation influence as a platform for classification and analysis, this study metaevaluates one organization's evaluations with an eye toward how those evaluations influenced it, whether through program improvement or other means. Metaevaluation proves a valuable means of exposing subtle forms of influence in the organization, as well as a way of revealing how it might evaluate in the future with an eye toward having even greater intentional influence.

KEYWORDS: *INGO evaluation; evaluation influence; humanitarian relief evaluation; program metaevaluation; intended use; unintended use*

Metaevaluation is potentially a valuable means of exploring how and in what ways evaluations influence the organizations and broader policy arenas in which they are conducted. Metaevaluation allows one to look back on the content and format of the evaluation, the timing of the evaluation and report, the environment in which the evaluation took place, and other details surrounding the evaluation to paint a picture of what factors may have contributed to the subsequent influence of the evaluation's process and findings over time. This metaevaluation of one international nongovernmental organization's humanitarian relief effort evaluations over a period of five years reveals how one approach to metaevaluation can highlight several ways in which an evaluation's reach can extend beyond the organization's direct use of the evaluation findings to include broader forms of influence inside and outside of the organization.

An evaluation has four potential functions: assessing the merit or worth of a program or policy, serving as a tool for compliance and oversight, contributing to the (research) knowledge base, or guiding the improvement of a policy or program (Mark, Henry, & Julnes, 2000). It follows, then, that metaevaluation

assesses how effectively an evaluation served one of these four functions. Summative metaevaluations, in particular, can highlight an evaluation's strengths and weaknesses (Stufflebeam, 2001), paving the way for better future evaluations. The metaevaluation described below, a summative metaevaluation, aimed to ascertain not only whether the evaluations examined effectively assessed merit and worth for program improvement, but whether they influenced the organization and beyond in other intended or unintended ways as well. Evaluation use, or utilization, has been prevalent in the evaluation literature for a few decades and refers to a change resulting from an evaluation or an evaluation report. Evaluators or organizations commissioning evaluations often place high priority on using findings for program improvement or, more broadly, to inform decision making. Use has been a central theoretical theme both for evaluation and for research-generated knowledge. However, organizations often do not have formal mechanisms for assessing whether their evaluations are used effectively or not, and virtually none have a systematic means for identifying indirect or unintended consequences of evaluation. In the case of nonprofits, this could be due in part to limited resources; moreover, those donating to nonprofits might prefer a more "direct" use of their money than metaevaluation activities. At its heart, use or utilization is an integral piece of the assessment of a program or policy's merit and worth.

The term "use," and its sister "utilization," have given way to the broader "influence" (Henry & Mark, 2003; Kirkhart, 2000) in more recent theoretical musings on evaluation. This shift reflects the limitations, semantic and otherwise, of the concept of use and opens the door for detecting unintended and subtle consequences of evaluating within an organization. Few organizations have studied formally the instance(s) of evaluation influence; perhaps this is in part because the factors leading to an evaluation's influence are poorly understood. Henry and Mark (2003) tackle this in examining the mechanisms that, for them, undergird the instances of influence. A better understanding of these mechanisms, and concrete examples, will go a long way toward helping evaluators to dissect how and under what circumstances evaluations are influential.

Metaevaluation allows for determining an evaluation's ultimate merit or worth by bringing to the surface the ways in which the evaluation did or did not influence the organization and its broader context. If one of the functions of an evaluation is to improve a program or policy, a core function of metaevaluation is to improve evaluation quality (Lipsey, 2000). In using metaevaluation to examine the extent of evaluation influence, this study considers the case of CARE, an international nongovernmental agency (INGO) whose emergency response work is well-recognized. CARE's humanitarian assistance work is a compelling single-case study because it is both representative of typical INGO work in this area and unique in its particular structure and dynamic. Virtually all of the evaluations considered in this study were of responses to emergencies to which other prominent INGOs also responded. By analyzing the case of CARE, this research purports to offer a window on the evaluation characteristics and dynamics typical of INGOs with a similar degree of reach and exposure.

CARE, as one of the eight largest and farthest-reaching agencies involved in humanitarian assistance (Cooley & Ron, 2002), has been a major recipient of the increased flow of funding to disaster relief efforts, both from USAID and from private individuals and foundations. As a signatory of the Red Cross Code of Conduct and an active member of Sphere, the Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP), and the Humanitarian Accountability Partnership (HAP), CARE is a forcible presence in the multitude of initiatives to harness humanitarian aid organization accountability.

CARE's current emergency response efforts extend from supplying food and water to providing shelter to facilitation of healthcare provision and delivery of essential supplies.

The study draws from two main data sources. The first of these data sources is all of the available evaluation reports from CARE's emergency response activity from 2000-05. These reports range from a brief summary to a multidocument behemoth. CARE uses four main formats for evaluation of emergency response efforts. (1) The Real Time Evaluation occurs in the middle of an intervention and assesses the success of the effort so far. (2) The After Action Review occurs just after an intervention and is typically a reflection session lasting three or four days and involving the staff members, temporary and permanent, who comprised the emergency response team. (3) The Final Evaluation occurs after the intervention and formally formulates the lessons CARE hopes to take away from the experience of the response for the future. (4) The Multi-Agency Evaluation involves the major INGOs who collaborated to mount a response in a large-scale emergency. These evaluations assess not only the effectiveness of each respective INGO, but examine the collaboration and coordination among all of the INGOs.

The second source of data is a series of interviews with twenty-five people associated with the evaluation process, from evaluators to field workers to management team executives. These individuals offer insight into their own perceptions and experiences of whether and how the evaluations influence CARE and their practice. The interviews examine both the perspectives of those who conduct the evaluations and the perceptions of those who purportedly read and used the final evaluation reports.

The research draws upon Kirkhart's (2000) Integrated Theory of Influence, which considers evaluations using three different gauges: intention (intended or unintended), source (process or results), and time (immediate, end-

of-cycle, or long-term). These dimensions inform the interview questions about whether and how the evaluations from 2000-05 affected later practice and policy.

The central analysis for the evaluation and interview data employs Henry and Mark's (2003) "pathways" of evaluation influence as the basis for examining how an evaluation affects an INGO from start to finish. Like Kirkhart, Henry and Mark find use to be a limiting term, and they advocate for thinking about evaluation influence instead. Their work culls from social science theories to propose pathways of influence which help to categorize the different levels at which influence might occur: (1) the individual, (2) the interpersonal, or (3) the collective (see Figure 1). Henry and Mark's taxonomy offers a starting point for examining one organization's treatment of evaluations in the emergency response arena. The analysis will use these three levels of influence to track interview responses and evaluation report data in order to observe how the report process and findings influenced various levels of communication, if at all.

Henry and Mark's (2003) taxonomy proposes a sort of menu for identifying and categorizing examples of influence. They are drawing from multiple disciplines, so some of their influence categories are more likely to show up in a large, decentralized INGO such as CARE than others. For example, in the "individual" level of influence, the "attitude change" mechanism is likely to surface in a study of CARE because it is a mechanism that easily lends itself to a program (as opposed to a policy). Determining whether an individual's attitude shifted is entirely feasible with interview data. Conversely, the "salience" mechanism is more about policy-related issues than about programs and so is not as likely to emerge from the CARE study. Other mechanisms, such as "elaboration," are difficult to pinpoint with interview data. Finally, it is more likely that the study will reveal examples of the individual-level and interpersonal-level mechanisms than the

collective-level mechanisms. This is because it is easier and takes less time to effect change at a programmatic level than at a policy level. Moreover, one of the five one individuals who participated in preliminary interviews for the study for the study mentioned his own impression that evaluation reports stop short of having policy-level influence at CARE, in part because the culture of learning there does not leave room for evaluation data in executive team agendas.

Henry and Mark's (2003) taxonomy draws from bodies of literature in several social science disciplines. Each of their three levels of influence has under it a number of change processes representing what evaluation influence could look like in any given context. As the CARE example demonstrates, the evaluator or the researcher may select from these change processes, or add to them, in order to cater a theory of influence to a particular situation. Figure 1 depicts how Henry and Mark's levels of influence break down into levels and "menu" items.
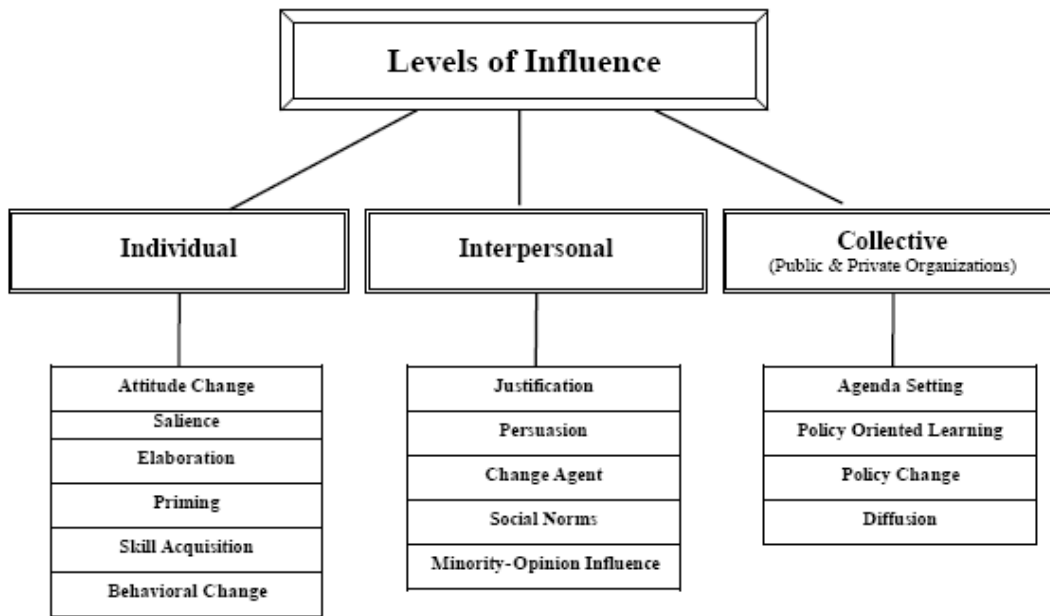

Figure 1. Mechanisms through Which Evaluation Produces Influences

Source: Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, *24*(3), p. 298.

The individual level concerns change brought about in a single person as a result of participating in an evaluation or reading the findings in an evaluation report. The types of influence for the individual level range from attitude change about an issue or program to actual behavioral change. The interpersonal level addresses types of influence occurring between two or more persons, as when one person uses an evaluation's findings to persuade another of his or her position. The collective level looks at change brought about at the organizational or interorganizational level, as when an evaluation's findings diffuse to another setting and foster change there.

Monica LaBelle Oliver

Methodology for the Metaevaluation and Analysis

Henry and Mark's (2003) taxonomy, described above, served as the basis for a coding scheme and checklist with which the researcher treated the interview and document data. Each level and subcategory in the taxonomy represented a different code, and decision-making criteria were assigned to each code. For example, for the Individual category/Skill Acquisition subcategory, coded IND_SA, the decision-making criterion was any evidence of "newly learned or honed skills." One example that fit this category was the participants' learning in an After Action Review evaluation to perform a SWOT (strengths, weaknesses, opportunities, and threats) analysis of their work to date. Any data falling into more than one category in the taxonomy was attributed to both categories. The researcher then paralleled the coded document data and the coded interview responses to identify intersections in the two sources that pointed to concrete examples of Henry and Mark's levels of influence. The evaluation reports were coded for their potential use or influence; in particular, the "findings" or "lessons-learned" sections of the evaluation reports showed where the evaluation might be used or have influence within the organization if the recommendations were taken to heart. The report contents were matched with the interview data as a way of identifying whether the interviewees remembered or even used evaluation report information.

Transcripts from five preliminary interviews suggested that more influence occurs at the individual and interpersonal levels than at the collective level. Indeed, the analysis that follows concludes that it is easier to identify examples of influence at the individual level and, to an extent, at the interpersonal level than it is at the collective level. Within these three levels, forms of process-related influence are more relevant to some CARE types of evaluations, whereas evaluation findings, end results, are more pertinent to other forms.

Findings and Discussion

The research brought examples of individual attitude and behavioral change to the surface, suggesting that evaluations do indeed have some influence in the organization on an individual level. There were no examples in the data of influence occurring at an interpersonal level. This is also true of the collective level, though the researcher acknowledges that CARE's complex organizational structure complicates the attempt to pinpoint just where the collective level begins. The scope and design of the research did limit the findings; though the researcher culled the organization's new evaluation standards, policy memos, and the like, in addition to the report content analysis and the interviews, the majority of the data for the research came from interviews with individuals, so it stands to reason that the Individual category was easier to identify in the data than were the other two categories.

Henry and Mark's (2003) taxonomy proved to be a useable tool for conducting a metaevaluation with the intent of examining the reach of influence of an organization's evaluation(s). In the case of the interpersonal and collective levels of influence, their taxonomy might be better suited to policy evaluation analysis than to program evaluation analysis. Nevertheless, their proposed levels of influence were germane to the CARE case and helped to reveal ways in which the organization might not have known that its evaluations had influence. For example, it was clear from the interview data that CARE employees cared whether they were included in the evaluation, whether as interviewees or as evaluation designers or otherwise. Their being included in the evaluation proceedings led them to be interested in the eventual

findings and more inclined to look to those findings as a legitimate source of information for their practice. Being excluded from the evaluation had the effect of making individuals less inclined to care about the subsequent findings. This sort of information can be valuable to CARE and similar organizations as they plan and design evaluations for the future, and might not have surfaced without meta-evaluating the evaluation report and processes.

Henry and Mark's (2003) evaluation influence taxonomy is but one example of a tool that can be adapted to the metaevaluation context. Their levels of influence provided a means for CARE to categorize its evaluation findings and employee perspectives in a meaningful way, allowing for a deeper understanding of evaluation's reach within the organization and providing a glimpse into how the organization and others like it might shape future evaluations in order to harness the positive benefits of that influence. Metaevaluation is a valuable tool for the deliberately introspective organization.

References

Cooley, A., & Ron, J. (2002). The NGO scramble: Organizational insecurity and the political economy of transnational action. *International Security*, *27*(1), 34.

Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, *24*(3), 293-314.

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation, No. 88* (pp. 5-23). San Francisco: Jossey-Bass.

Lipsey, M. (2000). Meta-analysis and the learning curve in evaluation practice. *American Journal of Evaluation*, *21*(2), 207-213.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving public and nonprofit policies and programs*. San Francisco: Jossey-Bass.

Stufflebeam, D. L. (2001). The meta-evaluation imperative. *American Journal of Evaluation*, 22(2), 183-209.