Commentary: Minimizing Evaluation Misuse as Principled Practice

J. BRADLEY COUSINS

Ethical Challenges, in my experience, is invariably interesting, often instructive and sometimes amusing. Some of the most engaging stimulus scenarios raise thorny evaluation practice issues that ultimately lead to disparate points of view about the nature of the issue and how to handle it (Datta, 2002; Smith, 2002). Despite my poor performance at the tables at the 2003 AEA conference in Reno, I am willing to gamble that the present “ethical challenge” does not fall into the “contentious case” category. I would wager that virtually all of my colleagues would agree: There is really no choice for our fledgling evaluator to make in this situation. She simply cannot proceed as if the problems with the data were minor and could be cloaked to no noticeable consequence. I hasten to add that this is an easy conclusion to reach within the comfort of a hypothetical scenario but an extremely difficult decision for anyone faced with the prospect of returning to a staple of Kraft Macaroni and Cheese Dinners after having lost that long-sought-after job in his or her desired field. And all this, after having slugged it out for so long in graduate school and having invested so heavily in tuition vouchers. Easier said than done, to be sure, yet principled decisions must prevail. The temptation to give in to the allure of personal comfort or questionable employer demands must be resisted.

To treat this as a case of “design degradation” that is “a bit worse than most” would be wrong professionally, morally, technically, and probably in several other ways. But concluding what the novice evaluator should not do says precious little about what she should do in this circumstance. Before considering that issue, I want to focus on the client’s, or “intended user’s” perspective. Let’s consider a framework articulating evaluation misuse as a possible source of professional practical guidance. Patton has long argued that fostering intended use by intended users is the evaluator’s responsibility. He also proposes that: As use increases so will misuse; policing misuse is beyond the evaluator’s control; and misuse can be intentional or unintentional (Patton, 1988, 1997). Thinking about evaluation misuse, I would offer, has the potential to help the evaluator understand standards of evaluation practice at concrete levels because it can aid in grasping the potential consequences for intended users of her choices and actions. Most discussion of evaluation ethics centers on the actions and behaviors of the evaluator (Christie & Alkin, 1999), yet it is my contention that such considerations are incomplete without serious attention to and analysis of the potential effects of such actions on
the user. It is what the user chooses to do with evaluation findings that ultimately affect a wide range of others, not the least of whom would be intended program beneficiaries.

Luminaries in evaluation theory such as Alkin, Patton and others have discussed evaluation misuse: How to define it? Identify or observe it? Track it? (see, e.g., Alkin, 1990; Alkin & Coyle, 1988; Christie & Alkin, 1999; Patton, 1988). My own distillation of some of these ideas is captured in Figure 1. It is divided into four quadrants by two orthogonal, dichotomous dimensions corresponding to the intended user’s choices and actions: (1) Use: evaluation findings are either used or not used, and (2) Misuse: evaluation findings are either handled in an appropriate and justified manner or they are not. Patton (1988), who initially proposed these ideas, used the terms misutilization and non-misutilization in describing the latter dimension. In the first quadrant we have the best case scenario, ideal use, where evaluation findings are employed in legitimate ways to support discrete decisions (instrumental use), foster learning (conceptual use), or symbolize or persuade (political use). These are familiar outcomes and, I would argue, we know a great deal about them and how to bring them about. I don’t think we can say the same about the remaining categories (Shulha & Cousins, 1997), although arguably our knowledge about Quadrant 4 has progressed notably through the study of non-use.

The undesirable consequences located in Quadrant 2 can manifest themselves in different ways. First, an intended user not well versed in evaluation logic and the methods of human science might inadvertently use bad data (mistaken use) by accepting it uncritically, taking

![Figure 1. Use and misuse of evaluation findings.](chart)
the evaluator’s word for it, and the like. In this instance, user intentions might be virtuous and supportable, yet the data are bad, possibly as a result of what Alkin calls misevaluation (Alkin & Coyle, 1988; Christie & Alkin, 1999). Misevaluation can arise from a host of sources including evaluator incompetence, evaluator mischief, unforeseen circumstances leading to design degradation, and the like. Alternatively, the intended user may be up to no good. Driven by motives of personal gain, hidden organizational agendas, or such, he may act to distort, fudge or otherwise misrepresent the findings (mischievous use). Mistaken use and mischievous use would be instances of misuse and their potential is worthy of analysis and consideration by our new evaluator in making her own choices (I will return to that point shortly).

Quadrants 3 and 4 in the framework come into play if evaluation findings are, in fact, not used. In Quadrant 3, supportable, valid findings somehow find themselves swept under the rug (abuse), again likely due to less than honorable, mischievous user intentions. Alkin calls this abuse or blatant non-use of sound evaluation findings (Alkin & Coyle, 1988; Christie & Alkin, 1999). Or, in the case of Quadrant 4, the user consciously decides not to use the findings because, for example, he or she knows they are flawed by virtue of evaluator incompetence, mischief, design degradation, or some such reason (rational non-use). This type of justified non-use may also arise where evaluation findings are withheld from the user for valid and supportable reasons, the evaluator knowing the data are bad being one such instance. In this case the intended user would be unaware of the existence of the findings, and, for obvious reasons, would not be in a position to use them. Rational non-use might also arise from untimely reporting, poor communication quality, unresponsive evaluation designs leading to irrelevant findings, and the like. It is equally important to note that the intended user may choose not to use defensible evaluation findings in light of competing information, changes in the decision or policy context, and a host of other factors that have been studied and documented extensively (political non-use) (see, e.g., Weiss, 1988, 1991). An interesting question would be: What are the criteria used to determine when a decision not to use sound evaluation findings is justified and when it is not? Some would argue that blatant non-use is entirely defensible in the political context of program and policy decision making. This issue reflects the sometimes ambiguous nature of misutilization and the political use of evaluation, a concern we have raised elsewhere (Shulha & Cousins, 1997) and one that Henry and Mark (2003) rightly suggest is problematic.

Now back to our hypothetical, shiny new evaluator. If a choice were on the table it would be to report and disseminate the evaluation findings, minimize obvious clues as to their inaccuracy and ignore the potential program consequences of this action OR to take a stand and develop and implement a principled plan of action.2 Let’s consider the consequences of the first choice. Assuming honorable intent on his behalf, the likely outcome for the intended user would be located squarely in Quadrant 2 under the flag of mistaken use. This would be a clear case of misevaluation: A competent, albeit inexperienced evaluator manipulating the evaluation data or their presentation so as to cover-up the obvious flaws. The unsuspecting intended user might integrate these distorted representations into program decision making and problem solving. If one thing can be said about bad data, they have the potential to lead to horrific decisions. We don’t know what sort of program is being evaluated in this case. However, if it is similar to the typical human-, social-, or educational-service interventions encountered by evaluators, then the lives, livelihood, or well-being of innocent, intended program beneficiaries could be on the line, not to mention risk to other program stakeholders in the form of injured self-esteem, damaged reputations or stunted careers (House, 1995). Another thing that can be said about bad data: When misevaluation is operative, any use of the resulting flawed findings constitutes misuse (Christie & Alkin, 1999).
On the other hand, the client might seek a second opinion and discover the flawed nature of the data and the cover-up. In that scenario, the user outcome might well be Quadrant 4 under the flag of rational non-use. Thankfully, intended program beneficiaries would be spared the ill-effects of potentially bad program decisions arising from bad data, but the consequences for the research firm and our rookie evaluator would be severe. “Your reputation is everything in this business!” is the cliché that seems to jump off the menu.

A final possible outcome would be associated with user ill-intent. In that case, evaluator choices would matter little because the user is likely to mold the findings to suit his purposes regardless of their veracity. Patton (1997) would argue, however, that the evaluator, in the course of developing relationships with intended users – as would be the case in utilization focused evaluation (UFE) – would have considerable control in preventing this sort of scenario. This might be through the early detection of ill-intended projected uses of the evaluation (and an evaluator bail-out option if need be) or later on as the intended user develops a sense of ownership of the evaluation by virtue of his proximity to and involvement with it. In Patton’s experience, “[intended users] become the champions of appropriate use, the guardians against misuse, and the defenders of the evaluation’s credibility when misuse occurs” (1997, p. 361).

What can and should be done in this scenario? I would propose a three-step plan that has as its goal moving from Quadrant 2 to Quadrant 1 via a temporary stop at Quadrant 4; that is, the evaluator working to maximize intended use, with the intended user having taken steps to minimize (or eliminate) misuse. Within this framework I will reference relevant guidelines from the Guiding Principles for Evaluators (American Evaluation Association, 1995). In doing so, I seek practical professional guidance by considering standards of evaluation practice while taking into account intended uses of the evaluation and the potential for its misuse.

Step 1: Conduct a thorough review and analysis of the evaluation (a meta-evaluation) and thoroughly document its flaws (as recommended by the Guiding Principle of Systematic Inquiry, which emphasizes adherence to high technical standards [A-1], and communicating accurately [A-3]). Sorely needed is a compelling case as to why the evaluation should not stay the current course. We need a solid read on misevaluation at this point.

Step 2: Confront the lead evaluator with this documented analysis and some ideas about alternative strategies. Assuming differential practical knowledge and experience, access to power, and stakes in the evaluation, the novice evaluator’s task of persuading the lead evaluator to alter the course in mid-evaluation is not likely to be an easy one. She has at her disposal the documented analysis, as well as the benefit of a wealth of resources and knowledge of evaluation logic arising from doctoral work (not to mention the AEA Guiding Principles and other standards of professional practice). This is where she must show some proverbial “backbone.” My sense is that if the lead evaluator is unresponsive to the arguments, the novice should confer with other, more senior members of the firm. Even if the lead evaluator is sympathetic, a team strategy seems warranted. The firm, after all, has a significant stake in this evaluation.

Step 3: Communicate the shortcomings of the evaluation to the client (assuming that the lead evaluator has been persuaded to provide the green light and/or such an action was negotiated in-house with reasonable senior members of the research firm) and ensure that the client understands thoroughly why proceeding as originally intended and planned would be highly problematic and inappropriate (suggested by the Integrity/Honesty Principle, which stresses the avoidance of misrepresentation [C-5] and the communication of concerns to the client [C-6]). If possible, identify parts of the evaluation that...
can be salvaged and offer these in light of the client’s decision making and problem solving needs. Having identified precisely what needs are not being met, negotiate a secondary plan that might involve reframing the evaluation, redesigning it, and/or gathering more and better data (consistent with Systematic Inquiry’s call to review planned evaluation questions and strategies [A-2], Integrity/Honesty’s support for identifying plural interests [C-3], and the recommendation of Responsibilities for General and Public Welfare to consider the full range of interests in planning [E-1] and explore broad assumptions, implications and potential side effects [E-2]). This plan is likely to be comparatively costly and the research firm would have to accept significant responsibility for the added expenditure (supported by the Honesty/Integrity Principle, which requires honest negotiations [C-1]). After all, the evaluator, as a representative of the firm, is now operating in damage control mode.

This approach will likely result in detrimental effects for the research firm and may ultimately lead to its dismissal by the client or, in a worst case scenario, civil litigation for contract violation or failure to deliver promised services. Presumably the firm would have legal counsel and would benefit from seeking legal advice before confronting the client. In any case, although its competency in the evaluation domain might be called into question, its integrity ultimately would not be. But this all assumes that Step 3 is in the cards. Another way the scenario could play out would be the immediate and uncelebrated dismissal of the “freshly minted,” yet now somewhat weathered, novice evaluator. That would be an unfortunate outcome. (Equally unfortunate is that there would have been little chance of her avoiding this outcome while maintaining professional integrity through, for example, a more thorough analysis of the firm before signing an agreement for gainful employment.) As non-trivial consolation, our evaluator’s professional integrity, pride, and dignity would survive the day—and anyway, Kraft Dinners taste okay.

In closing, let’s return to the prospect of integrating evaluation use (and misuse) theory with standards of professional practice as a recommended course of action. Recent theoretical developments have cast evaluation use theory as being ‘old school,’ arguing instead in favor of a more broadly conceived theory of influence (Henry & Mark, 2003; Kirkhart, 2000; Mark & Henry, 2004). I think these developments are exciting and they carry with them the potential to stimulate healthy scholarly debate and advance our field. Kirkhart’s (2000) reconceptualization of use has provided Henry and Mark with a platform from which to add considerable value in terms of implications for thinking about and studying patterns of evaluation impact. Entering into a debate on the merits of this new direction lies outside my present purpose, but I would like to identify one aspect of Henry and Mark’s argument that is relevant to the current discussion. Among their multifaceted rationales for moving “beyonf use” is the observation that “use is inadequate as a motivation for evaluators, because it fails to provide a moral compass (House, 1995) and falls short of recognizing and promoting the ultimate purpose of evaluation as social betterment” (Henry & Mark, 2003, p. 295). The authors provide several other justifications for their case, but for the present purpose, I will restrict my focus and make two points.

First, in and of itself, a theory of influence (Kirkhart, 2000), similar to a theory of use, does not have a built-in moral compass. Henry and Mark have chosen to introduce such a moral compass to a theory of influence by proclaiming social betterment to be the ultimate purpose of evaluation. Whether all evaluation should be guided by this or some other ideology is a matter for debate and discussion elsewhere. For now, I would observe that the utilization-oriented
evaluators known to me would not think twice about turning down an evaluation contract for any number of good professional reasons, many of which are eloquently documented by Smith (1998). This is no accident. These individuals, I would offer, are acting in accord with Patton’s (1997) counsel: They are being clear about and guided by their own morals; they are motivated to understand and be sensitive to whose interests are being represented in the evaluation; and they are making informed and careful choices concerning with whom to work. In short, though motivated to produce useful evaluations they are working within a framework of professional standards (e.g., the Guiding Principles for Evaluators) and as such, I would argue, are well equipped to maintain solid footing on the moral high ground.

Mark and Henry (2004) are drawn to a theory of influence because it enables one to consider and examine underlying change mechanisms that lead to evaluation influence and impact whereas, in their view, evaluation use theory is deficient in this regard. This is an interesting assertion and an undeniable stimulus for further inquiry. For now, however, the second point I would propose is that by virtue of engaging in a conceptual analysis of the potential for evaluation misuse – one guided by a framework such as that appearing in Figure 1 – our new evaluator can make relatively concrete for herself the kinds of effects on the intended user (positive and negative) her choices and behaviors might have. Such analysis does so in ways that standards of professional practice (e.g., the Guiding Principles), on their own, do not, and that a theory of influence has yet to address in a direct way. On the latter point, Mark and Henry (2004) provide an elaborate and well developed menu of mechanisms through which evaluation influence may be tracked, but as of yet they have little to say about how, when, and why such influences may be less than desirable (a theory of misinfluence, perhaps?). Having as a guide the prospect of achieving social betterment through evaluation may be virtuous but this, by itself, does not provide evaluators with much in the way of concrete practical advice in the short term. Thinking about evaluation use and misuse within a framework of professional standards of practice, on the other hand, can offer such help. Though the issue needs to be studied much more thoroughly (Shulha & Cousins, 1997), I believe that through appropriate actions and relationship building evaluators can play a strong role in steering intended user outcomes toward ethical, legitimate use.

ACKNOWLEDGMENTS

This paper benefited significantly from the insights and wisdom of two colleagues who remain anonymous.

NOTES

1. Although, for convenience, I use the term “intended user” in the singular, I acknowledge that program and policy decision makers rarely act alone. I tend to use “evaluator” in the singular for the same reason.

2. Other intermediate alternatives exist, such as petitioning to be reassigned to a different project, or seeking alternative employment. For the present purposes, however, let’s rule out avoidance strategies and consider the choice to be binary.

3. My understanding is that Henry and Mark own this claim. My reading of House (1995) reveals no mention of use as a problematic motivation for evaluators, or reference to navigational instruments, moral or otherwise. However, House (quoted in Alkin, 1990, pp. 244–246, and cited by Henry and Mark
later in their article) maintains that unchecked catering to client interests is immoral. He later refers to this as the ethical fallacy of "clientism" (House, 1995).

4. I use the phrase “utilization-oriented evaluators” to depict those motivated by the prospect of fostering the use of evaluation for decision making and problem solving. This would include but not be limited to Patton’s UFE approach (1997). See Alkin and Christie (2004) for an extended summary and analysis of the many and varied utilization-oriented approaches to evaluation.

REFERENCES


