

Internal and external evaluators experience the same ethical dilemmas, and both work toward resolving them in principled, justified ways. The different communities they occupy, however, affect how they are likely to resolve their ethical dilemmas.

Rights, Responsibilities, and Duties: A Comparison of Ethics for Internal and External Evaluators

Sandra Mathison

All evaluators probably try to adhere to a common set of ethical principles, such as the American Evaluation Association's *Guiding Principles for Evaluators* or *The Program Evaluation Standards*, and both internal and external evaluators are as likely (or unlikely) to engage in ethical evaluation practice. Although many discussions of internal-external evaluation revolve around questions about objectivity, bias, and credibility, I will adopt the assumption (based primarily on a lack of evidence in the literature to indicate otherwise) that neither internal nor external evaluators are inherently more or less objective, biased, or credible. Therefore, the differences that will be highlighted in this chapter are issues related to the ethics of the situation (evaluating from inside or outside) and the attention evaluators pay to the ethics of the evaluand. Most determinations of what is a right thing to do or a right way to be are made situationally, and the situations for internal and external evaluators are different. Internal and external evaluators occupy different "communities" and it is the nature of those communities that is critical for considering the resolution of ethical dilemmas.

The first section of the chapter will examine the question of whether the ethical dilemmas faced by internal and external evaluators are themselves different. The second section will examine the details of situational differences based on the conclusion from the previous section that, in fact, most, if not all, ethical dilemmas are common to all evaluators. I will argue that given a backdrop of common ethical principles for evaluation practice, a community framework helps us to understand the differences between internal and external

evaluators. In writing this chapter I have relied on existing research on internal-external evaluation and ethics, although it should be noted that the research base is limited. So this chapter should also be considered an attempt to take what we do know about ethics and the internal-external distinction and build a theoretical framework from that, which itself requires further examination and empirical testing.

Are Ethical Dilemmas Different for Internal and External Evaluators?

The short answer to this question is no. There seems to be no concrete evidence in the literature to support the contention that the nature of ethical dilemmas differs depending on whether one is an internal or external evaluator. Although internal evaluators are less likely to report encountering ethical challenges, when they do, these challenges are not substantially different in kind from those reported by external evaluators (Morris and Cohn, 1993). Many ethical dilemmas stem from conflicting values over goals, processes, outcomes (Smith, 1985), and theoretical frameworks (Kirkhart, 1985), but these dimensions do not distinguish internal from external evaluators.

Still, there is a persistent view that ethical challenges for internal evaluators are unique. Adams (1985) suggests internal evaluators might encounter (1) pressure to downplay negative and emphasize positive findings, (2) valuing of routine, nonthreatening evaluation activities, (3) interest in having an evaluation unit but not in using evaluation results, or (4) reduced access to privileged information. Internal evaluations are suspect because, it is presumed, they can be manipulated more easily by administrators to justify decisions or pressured to present positive findings for public relations purposes (House, 1986). Some suggest that these latter two problems may be reduced if the client for the internal evaluation is sufficiently distant from both the evaluand and the evaluator within the organization, that is, is high enough in the hierarchy such that the "truth" is what matters most (Worthen, Sanders, and Fitzpatrick, 1997; see especially Chapter Twelve).

External evaluators also work with clients who may make an undue attempt to shape the design, execution, and reporting of the evaluation. Scriven (1983) posits a managerial ideology that describes, in general, the propensity for evaluators and evaluations to attend disproportionately to the interests and needs of program managers. This is not more or less the case for internal or external evaluators, although there is evidence that internal evaluators typically respond to decision makers' needs and try to increase the usability of evaluation information by decision makers (Love, 1983; Mathison, 1991a; Patton, 1997; Reavy, Littell, Gonda, and O'Neill, 1993; Winberg, 1991). However, many external evaluators operate within this ideology as well.

Additionally, symbolic uses of evaluation are not limited to either external or internal evaluations, although the symbolic nature of the evaluation might be more apparent to internal evaluators. For example, internal evalua-

tors may be directly asked for public relations information to support a program, but external evaluators are often employed to obtain information to be used in seeking an extension of or application for funding. All programs have the potential for ethical or unethical practices, and all clients and evaluations done for them have the potential for ethical and unethical practices. And both internal and external evaluators serve masters who control their fates in some way, to some degree.

Perhaps an illustration of precisely the same ethical dilemma I have encountered as both an internal and external evaluator will illustrate the commonality of such dilemmas. In both roles I have been involved in evaluations of teacher development projects, projects frequently of modest duration and likely to have limited, specific impact on the teachers involved. In both situations I have been pressured (even ordered) to examine student outcomes from these projects. Politically and ideologically, to establish positive student outcomes is perceived to be the best evidence that teacher development was successful. In both the internal and external evaluation, this demand was unwarranted, given the limited nature of the interventions and the conceptualizations underlying them. This one example is not meant as a definitive counter-example, but it is typical of my experience as an evaluator.

The difference lies, then, not in the dilemmas, but in the ways in which they are resolved, which will be taken up in the next section. Internal and external evaluators do work under different conditions, operate within different communities, and therefore approach ethical dilemmas in potentially different ways.

Understanding the Situational Differences of Internal and External Evaluators

Internal and external evaluators occupy different communities in their work as evaluators. Figure 3.1 situates internal and external evaluators with respect to organizations, the larger discipline or product context within which organizations operate, and the evaluation profession. Each of the spheres can be thought of as a community, not in the sense of place, but in the sense of “a shared commitment to a set of core values” (Etzioni, 1996, p. 13), or what Dewey referred to as “a mode of associated living, of conjoint communicated experience” (1916, p. 87). Communities are made up of individuals (like evaluators) exercising individual rights, responsibilities, and interests, as well as the rights, responsibilities, and interests of the community as a whole. For all evaluators, ethical behavior is a balance of personal integrity, personal interest, and community interest. As will be discussed further, internal evaluators have fewer communities within which they operate, but they are very closely tied to those communities. External evaluators have a wider array of communities within which they function, and although they no doubt have close ties with one (or more) of those communities it is less likely to be the community or organization within which the program they evaluate is situated. Neither of

these circumstances has moral, political, or methodological superiority; they are simply different.

Internal evaluators work within a particular organization that is itself embedded in a larger context that is composed of similar and related organizations. So, for example, an internal evaluator might be employed by the Rivet and Staple Company, which is one of many organizations that does light manufacturing. Or an internal evaluator might be employed by the Caring and Giving Foundation, which is part of the larger context of not-for-profit charitable foundations.

External evaluators, however, may work across a number of organizations and perhaps across a number of different discipline or product contexts as well. For example, an external evaluator might be a faculty member at Big Time University and simultaneously be doing an evaluation study for the Neighborhood Elementary School, which is part of the larger context of American public schools, and an evaluation for the Boys and Girls Club, which is part of the larger context of not-for-profit social-service delivery systems. Both internal and external evaluators are part of the evaluation profession evidenced by their membership in professional associations, subscriptions to professional journals, and so on.

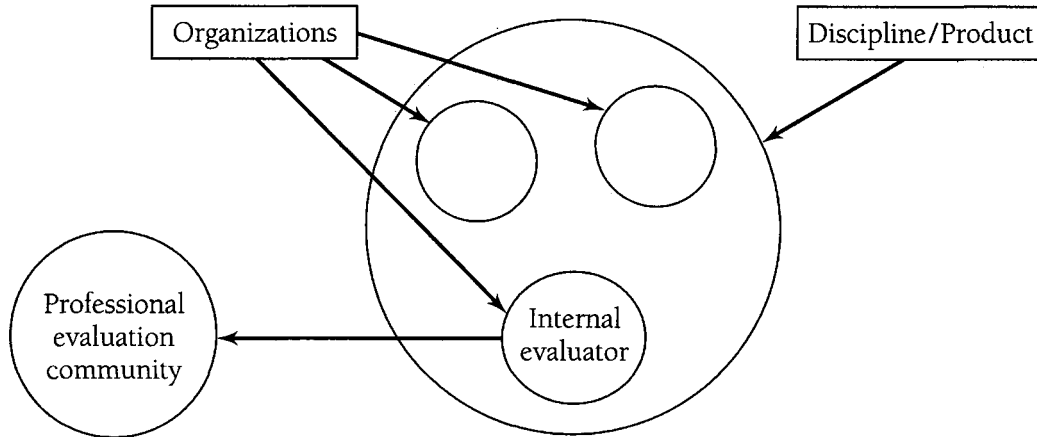
A critical difference between internal and external evaluators is this definition of community; internal evaluators are more oriented to communities consistent with their workplaces, whereas external evaluators are oriented to a diverse array of communities, including, but not limited to, the professional evaluation community, their academic discipline, employers (such as universities or consulting firms), funding agencies, local business communities, and so on. For internal evaluators, this community is defined first and foremost by the *organization* within which she is working, as well as the discipline or product context of which that organization is part. The evaluation profession is, then, just one piece of this mix. The internal evaluator has a role to play in each of these areas, roles that may conflict with one another (Mathison, 1991b). So the community for an internal evaluator begins with the organization and moves out to the discipline or product community and the evaluation profession. The ethical dilemmas encountered by an internal evaluator are situated specifically and particularly within the organization and involve people the evaluator knows and works with routinely. The most enduring element of the internal evaluator's community is the organization.

The community in which the external evaluator operates is not as clearly defined, because the particular organizations and even disciplines in which he works are changeable. The ethical dilemmas encountered by an external evaluator are situational, occur in a changing work environment, and involve people the evaluator may or may not know and does not work with routinely. The primary community for an external evaluator is less predictable and covers a wide range of possibilities, as indicated in Figure 3.1. Again, the professional evaluation community is a part of this mix.

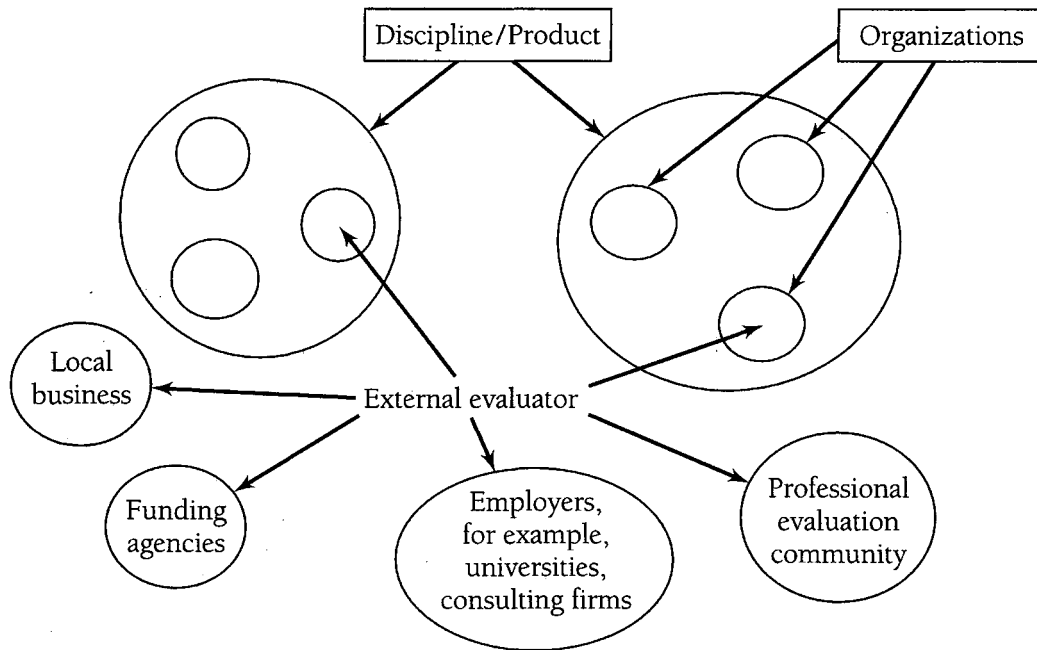
When internal or external evaluators encounter an ethical dilemma, the context in which they understand and then resolve the dilemma is, therefore,

Figure 3.1. A Comparison of Communities for Internal and External Evaluators

Communities for Internal Evaluators



Communities for External Evaluators



different. I return to the situation I described in the previous section—the demand by my boss or client that I assess student outcomes as part of the evaluation of teacher development projects—to illustrate these differences. As suggested, the occurrence of the dilemma is not related to the internal-external distinction, but I believe the resolution of the dilemma and the impact of that resolution does depend on it.

As an internal evaluator responsible for the evaluation of a teacher development project directed to primary grade teachers and intending to enhance their knowledge and skills in teaching mathematics, I was “enjoined” to assess students’ mathematics performance as a part of the evaluation. Given the nature of the intervention and the teachers’ serious lack of content and pedagogical knowledge of mathematics, such indicators would likely suggest program failure, whereas other indicators (such as changes in teachers’ knowledge about math, amount of time spent teaching math, variety of math activities, and so on) were more connected to the program staff’s intentions, better matched to program intensity, and hence more appropriate indicators of program success.

As an internal evaluator, one loses some of the aura of the external expert (Adams, 1985); resolving an ethical dilemma thus often requires what I would call an *action proof*. (I will contrast this with an *academic proof* for an external evaluator.) In other words, although the internal evaluator is presumed to have special expertise, it is not mysterious, and recommendations about what and how to do things probably work best if the evaluator can show her recommendation is sound. An action proof is essentially a “show-me” strategy (rather than a reasoning strategy) requiring a context-specific demonstration. In other words, convincing program staff that something is a right or wrong thing to do must be embedded in that particular context and involve a demonstration also embedded within that context.

In the above example, I suggested to program staff that we look at existing student performance data from the year previous to the project beginning and during the first year of the project to determine the usefulness of such data in evaluating the teacher development project. This demonstration of what the use of student outcome data would show in conjunction with the use of indicators more closely coupled with the program’s intents—in other words, showing what are good and not-so-good indicators of the program’s success—convinced program staff they might potentially draw unnecessary criticism by focusing on student outcomes rather than teacher outcomes in this evaluation.

This strategy was important within the context of this particular evaluation, but also important insofar as internal evaluators must consider the long-term and related consequences of their choices. Within this organization, such a strategy builds an understanding extending beyond the specific project and into a larger discourse about how to think about appropriate indicators of success. This approach also demonstrates what is probably a greater degree of conciliation on the part of internal evaluators—compromise is important in the creation of a community of people who share larger common goals.

As an external evaluator, I have also been enjoined to assess student performance as an indicator of the success of a teacher development project. In one case, a modest (although expensive) intervention to help elementary teachers use their own schoolyards to teach science, the program developers had promised their funding agency that the project would make a positive difference in student outcomes. As with many innovative curricular approaches, the more likely immediate effect of changes in pedagogy and content is diminished student performance, particularly if what one uses are standard school indicators of such performance. Because the program developers had promised the funding agency such a demonstration, there was significant pressure to use standard student outcome measures. Unlike the internal evaluation situation, providing an academic proof was sufficient. This proof relies more on the invocation of authority and expertise vested in me as an expert evaluator (who is being paid substantial amounts of money for good advice) and the use of other examples from the evaluation and research literature. So illustrations of the negative impact of educational innovations on standardized measures of student achievement were compelling, even though they were not specific to this project or even the particular focus of this project, that is, science.

Although an external evaluator could also use an action proof, there is less demand for it and less need for it, as the external evaluator is much less concerned about the consequences for this project and others within the organization—she may or may not be involved and so has less vested interest in creating contexts for specific kinds of discourse. (I have noted elsewhere that this is a potentially serious problem with external evaluation, and suggest a modified role for external evaluators whereby they would have long-term relationships with organizations, rather than being evaluators of specific activities or projects [Mathison, 1994].) Similarly, although external evaluators want to maintain positive relationships with their clients, they can afford to be less conciliatory than the internal evaluator, as their connections to the organization are circumscribed and particular. And they are afforded more latitude as an expert than is the internal evaluator.

The fundamental basis of these differences is based on the nature of the community to which internal and external evaluators belong. The community for internal evaluators is their workplace; regardless of the kind of organization, they are committed to making that community work well, both in particular instances and in the long term. Their decisions about how to respond to the myriad ethical dilemmas posed by evaluation are contextualized by this primary affiliation with their particular workplace. External evaluators, however, operate in less particularized communities and perhaps the greater unifying community may be that of other evaluators (although I am not suggesting that internal evaluators are not also connected to the community of professional evaluators). Their decisions about how to respond to ethical dilemmas are informed by general practices, other evaluation experience, demands of other work communities, and with reference to what other professional evaluators would do in this sort of instance.

It might be argued that internal evaluators, through their use of action proofs and familiarity with the organization, the program, and the stakeholders, have greater potential for a sustained impact on the ways in which evaluation can contribute to positive outcomes within an organization. (External evaluators might also consider using action proofs as well, even though often an academic proof is sufficient. This might be particularly helpful if an external evaluator adopts principles that support sustained organizational learning and change.) I will argue in the next section that internal evaluators may be less well situated to raise questions about the ethics of the evaluand itself than are external evaluators.

Judging the Ethics of the Evaluand

The previous section discussed ethical issues related to doing an evaluation well and how the internal-external distinction makes a difference. Evaluators are also meant to judge the ethics of the evaluand—to raise questions about whether the job is being done right and to raise questions about whether the right job is being done, what Hendricks (1985) described as a need to judge the appropriateness of services.

Let me return to my experiences as an internal evaluator for the mathematics project described earlier. This project (actually many projects) adopted an underlying principle that the way to improve schools was to employ a standard research, development, and diffusion model. This model focuses on developing packaged curricular materials and teacher education that can be widely disseminated regardless of the circumstances in which it is implemented. The lack of success of this model is well documented (House, 1981; Mathison, 1990). As the internal evaluator, this approach to educational change seemed to me misguided (given its demonstrated lack of success), equivalent to not doing the right thing. The project might be more successful at real, sustained educational change if it focused on specific schools, working from a grassroots action research model, rather than an expertise-driven research model. As an internal evaluator, it was difficult to raise questions about these fundamental principles within the projects. My expertise was presumed to be in evaluation, understood as expertise in technique, so my reflections on curricular and pedagogical foundations were considered to be less informed than reflections on data analysis strategies, for example. Little that I did was a compelling challenge to the conceptualization of the project, never mind effective in altering these foundational principles.

This is an innocuous example when compared to, for example, addressing issues such as racism, sexism, or classism that may be inherent in a particular program or project. Here, the internal evaluator is often in a weak position to confront these matters, especially if they are rampant within or condoned by the organization. Fear of loss of one's job, friendships, or dissolution of one's community or organization are serious risks for the internal evaluator. The internal evaluator as whistle-blower entails substantial personal and organizational

risks. And raising questions about the ethical stance of a project or activity within the organization potentially puts the internal evaluator in a position of questioning her own complicity with unethical practices. External evaluators, while experiencing the above constraints to a certain degree, are in a better position to voice serious concerns, in large part because they do not belong to these organizational or workplace communities and do not need to sustain relationships and commitments to doing a good thing within that context.

For example, as an external evaluator working with a group of doctoral students to evaluate an after-school program for inner-city teens, it was much easier for us to raise issues of racism that were subtle and insidious within the program. The program focused on a transition-to-work idea and, as such, the teens were responsible for finding their own internships within the organizations, in this case a museum. This organization provided a vast array of potentially valuable experiences for them, but for the most part they either floundered about looking for work or worked in menial jobs, such as cleaning in the cafeteria. Underlying the program's inability to provide positive, meaningful work experiences for these teens was a deficit model of inner-city, African-American teens—one that presumed they weren't very able, by definition had psychological problems, and couldn't really be counted on in things that matter. The program itself was ethically flawed. As external evaluators we could much more easily raise these issues of racism than could an internal evaluator. We were not seen as having a vested interest (although the staff did feel betrayed by us for raising these issues, and the teens were grateful for being given an opportunity to reveal their perceptions) and were presumed by our expertise to be doing a fair job. Did the racism diminish as a consequence of the evaluation? I do not think so, but the issue at least got talked about by a range of program stakeholders. This is always the downside for external evaluators—they have the power and authority to raise such issues but little efficacy when it comes to ameliorating unethical programs. Naming unethical programs or practices does not, of course, guarantee their elimination.

Both internal and external evaluators should look at the ethics of the things they evaluate, and I presume they do. Internal evaluators are more at risk when raising such questions than are external evaluators, however.

Conclusion

Internal and external evaluators experience the same ethical dilemmas and both work toward resolving them in principled, justified ways. However, they occupy different communities, and these communities have an impact on how ethical dilemmas are addressed and what strategies are most effective in resolving them. In some situations internal evaluators have the edge (such as in creating communities of discourse about how to do evaluation well), whereas in other situations external evaluators do (such as in raising questions about unethical intentions or practices within programs). The purpose of this chapter is not to value one situation over the other but to suggest to both internal

and external evaluators that thinking about themselves as embedded within communities is one way to proceed in dealing with ethical dilemmas in their practice. Also, the analysis presented here is not meant to suggest that these community structures are fatal in determining how internal and external evaluators might approach ethical dilemmas, and indeed both groups have much to learn from each other.

References

- Adams, K. A. "Gamesmanship for internal evaluators: Knowing when to 'hold 'em' and when to 'fold 'em.'" *Evaluation and Program Planning*, 1985, 8, 53–57.
- Dewey, J. *Democracy and Education*. New York: The Free Press, 1916.
- Etzioni, A. *The New Golden Rule: Community and Morality in a Democratic Society*. New York: Basic Books, 1996.
- Hendricks, M. "Should Evaluators Judge Whether Services Are Appropriate?" *Evaluation and Program Planning*, 1985, 8, 34–44.
- House, E. R. "Three Perspectives on Innovation: Technological, Political, and Cultural." In R. Lehming and M. Kane (eds.), *Improving Schools*. Beverly Hills, Calif.: Sage, 1981.
- House, E. R. "In-House Reflection: Internal Evaluation." *Evaluation Practice*, 1986, 7 (1), 63–64.
- Kirkhart, K. "Analyzing Mental Health Evaluation: Moral and Ethical Dimensions." *Evaluation and Program Planning*, 1985, 8, 13–23.
- Love, A. J. (ed.). *Developing Effective Internal Evaluation*. New Directions for Program Evaluation, no. 20. San Francisco: Jossey-Bass, 1983.
- Mathison, S. "Reflections on Evaluating Innovative Curriculum Projects." In I. Wirsaup and R. Streit (eds.), *Developments in School Mathematics Education Around the World, Vol. 2*. Reston, Va.: National Council of Teachers of Mathematics, 1990.
- Mathison, S. "What Do We Know about Internal Evaluation?" *Evaluation and Program Planning*, 1991a, 14, 159–165.
- Mathison, S. "Role Conflicts for Internal Evaluators." *Evaluation and Program Planning*, 1991b, 14, 173–179.
- Mathison, S. "Rethinking the Evaluator Role: Partnerships Between Organizations and Evaluators." *Evaluation and Program Planning*, 1994, 17, 299–304.
- Morris, M., and Cohn, R. "Program evaluators and ethical challenges: A national survey." *Evaluation Review*, 1993, 17, 621–642.
- Patton, M. Q. *Utilization-Focused Evaluation: The New Century Text* (3rd ed). Thousand Oaks, Calif.: Sage, 1997.
- Reavy, P., Littel, L., Gonda, G., and O'Neill, I. "Evaluation as Management Support: The Role of the Evaluator." *Canadian Journal of Program Evaluation*, 1993, 8 (2), 95–104.
- Scriven, M. "Evaluation Ideologies." In G. F. Maddaus, M. Scriven, and D. L. Stufflebeam (eds.), *Evaluation Models*. Boston: Koeber-Nijoff, 1983.
- Smith, N. L. "Some Characteristics of Moral Problems in Evaluation Practice." *Evaluation and Program Planning*, 1985, 8, 5–11.
- Winberg, A. "Maximizing the Contribution of Internal Evaluation Units." *Evaluation and Program Planning*, 1991, 14, 167–172.
- Worthen, B., Sanders, J., and Fitzpatrick, J. *Program Evaluation: Alternative Approaches and Practical Guidelines*. New York: Longman, 1997.

SANDRA MATHISON is associate professor in the Department of Educational Theory and Practice at the University at Albany, State University of New York.