THE NATURE AND PRACTICE OF PROFESSIONAL JUDGMENT IN PROGRAM EVALUATION

Introduction

The Nature of the Evaluation Enterprise

What Is Good Evaluation Practice?
Methodological Considerations
Problems With Experimentation as a Criterion for Good Methodologies
The Importance of Causality: The Core of the Evaluation Enterprise
Alternative Perspectives on the Evaluation Enterprise
Reconciling Evaluation Theory With the Diversity of Practice
Working in the Swamp: The Real World of Evaluation Practice
Common Ground Between Program Evaluators and Program Managers
Situation Professional Judgment in Program Evaluation Practice

Acquiring Knowledge and Skills for Evaluation Practice

Professional Knowledge as Applied Theory
Professional Knowledge as Practical Know-How
Balancing Theoretical and Practical Knowledge in Professional Practice

Understanding Professional Judgment

A Modeling of the Professional Judgment Process
The Decision Environment
Values, Beliefs, and Expectations
Acquiring Professional Knowledge
Improving Professional Judgment in Evaluation
Through Reflective Practice 458
Guidelines for the Practitioner 458
The Range of Professional Judgment Skills 461
Ways of Improving Sound Professional Judgment
Through Education and Training-Related Activities 461
Teamwork and Professional Judgment 463
Evaluation as a Craft: Implications for Learning to
Become an Evaluation Practitioner 464

Ethics for Evaluation Practice 465
The Development of Ethics for Evaluation Practice 465
Ethical Evaluation Practice 467
Cultural Competence in Evaluation Practice 468

The Prospects for an Evaluation Profession 469

Summary 470

Discussion Questions 471

Appendix 473
Appendix A: Fiona’s Choice: An Ethical Dilemma for a Program Evaluator 473
Your Task 474

References 474
Chapter 12 begins by reflecting on what good evaluation methodology is, and reminding the reader that in the evaluation field there continues to be considerable disagreement around how we should design evaluations to assess program effectiveness and, in so doing, examine causes and effects. We then look at the diversity of evaluation practice, and how evaluators actually do their work. Developing the capacity to exercise sound professional judgment is key to becoming a competent evaluator.

Much of Chapter 12 is focused on what professional judgment is, how to cultivate sound professional judgment, and how evaluation education and training programs can build in opportunities to learn the practice of exercising professional judgment. Key to developing one’s own capacity to render sound professional judgments is learning how to be more reflective in one’s evaluation practice. We introduce evaluation ethics and connect ethics to professional judgment in evaluation practice.

Throughout this book, we have referred to the importance of professional judgment in the practice of evaluation. Our view is that evaluators rely on their professional judgment in all evaluation settings. Although most textbooks in the field, as well as most academic programs that prepare evaluators for careers as practitioners, do not make the acquisition or practice of sound professional judgment an explicit part of evaluator training, this does not change the fact that professional judgment is an integral part of our practice.

To ignore or minimize the importance of professional judgment suggests a scenario that has been described by Schön (1987) as follows:

In the varied topography of professional practice, there is the high, hard ground overlooking a swamp. On the high ground, manageable problems lend themselves to solutions through the application of research-based theory and technique. In the swampy lowland, messy, confusing problems defy technical solutions. . . . The practitioner must choose. Shall he remain on the high ground where he can solve relatively unimportant problems according to prevailing standards of rigor, or shall he descend to the swamp of important problems and non-rigorous inquiry? (p. 3)

Evaluation can be viewed as a structured process that creates and synthesizes information that is intended to reduce the level of uncertainty for stakeholders about a given program or policy. It is intended to answer questions (see the list of evaluation questions discussed in Chapter 1) or test hypotheses, the results of which are then incorporated into the information bases used by those who have a stake in the program or policy.
What Is Good Evaluation Practice?

Methodological Considerations

Views of evaluation research and practice, and in particular about what they ought to be, vary widely. At one end of the spectrum, advocates of a highly structured (typically quantitative) approach to evaluations tend to emphasize the use of research designs that ensure sufficient internal and statistical conclusions validity that the key causal relationships between the program and outcomes can be isolated and tested. According to this view, experimental designs are the benchmark of sound evaluation designs, and departures from this ideal are associated with problems that either require specifically designed (and usually complex) methodologies to resolve limitations, or are simply not resolvable—at least to a point where plausible threats to internal validity are controlled.

In the United States, evaluation policy for several major federal departments under the Bush administration (2001–2009) emphasized the importance of experimental research designs as the “gold standard” for program evaluations. As well, the Office of Management and Budget (OMB) reflected this view as it promulgated the use of the Program Assessment Rating Tool (PART) process between 2002 and 2009. In its 2004 guidance, the OMB states the following under the heading “What Constitutes Strong Evidence of a Program’s Effectiveness?” (OMB, 2004):

The revised PART guidance this year underscores the need for agencies to think about the most appropriate type of evaluation to demonstrate the effectiveness of their programs. As such, the guidance points to the randomized controlled trial (RCT) as an example of the best type of evaluation to demonstrate actual program impact. Yet, RCTs are not suitable for every program and generally can be employed only under very specific circumstances. (p. 1)

The No Child Left Behind Act (2002) had, as a central principle, the idea that a key criterion for the availability of federal funds for school projects should be that a reform has been found, through scientifically based research to significantly improve the academic achievement of students participating in such program as compared to students in schools who have not participated in such program; or . . . has been found to have strong evidence that such program will significantly improve the academic achievement of participating children. (Sec. 1606(a)11(A & B))

In Canada, a major federal department (Human Resources and Skills Development Canada) that funds evaluations of social service programs has specified in guidelines that randomized experiments are ideal for evaluations, but at minimum, evaluation designs must be based on quasi-experimental research designs that include comparison groups that permit before-and-after assessments of program effects between the program and the control groups (Human Resources Development Canada, 1998).
Problems With Experimentation as a Criterion for Good Methodologies

In the United States, privileging experimental and quasi-experimental designs for evaluations in the federal government has had a significant impact on the evaluation community. Although the “paradigm wars” were thought to have ended or at least been set aside in the 1990s (Patton, 1997), they were resurrected as U.S. government policies emphasizing the importance of “scientifically based research” were implemented. The merits of randomized controlled trials (RCTs) as the benchmark for high-quality evaluation designs have again been debated in conferences, evaluation journals, and Internet listserv discussions.

Continued disagreements among evaluators about the best or most appropriate ways of assessing program effectiveness will affect the likelihood that evaluation will emerge as a profession. Historically, advocates for experimental approaches have argued in part that the superiority of their position rests in the belief that sound, internally valid research designs obviate the need for the evaluator to “fill in the blanks” with information that is not gleaned directly from the (usually) quantitative comparisons built into the designs. The results of a good experimental design are said to be more valid and credible, and therefore more defensible as a basis for supporting decisions about a program or policy. Experimentation has also been linked to fostering learning cultures where new public policies are assessed incrementally and rationally. Donald Campbell (1991) was among the first to advocate for an “experimenting society.”

Although experimental evaluations continue to be important (Ford, Gyarmati, Foley, Tattrie, & Jimenez, 2003; Gustafson, 2003) and are central to both the Cochrane Collaboration (Higgins & Green, 2011) in health-related fields and the Campbell Collaboration (2010) in social program fields as an essential basis for supporting the systematic reviews that are the mainstay of these collaborations, the view that experiments are the “gold standard” does not dominate the whole evaluation field. The experiences with large-scale evaluations of social programs in the 1970s, when enthusiasm for experimental research designs was at its highest, suggested that implementing large-scale RCTs was problematical (Pawson & Tilley, 1997).

Social experiments tended to be complex and were often controversial as evaluations. The Kansas City Preventive Patrol Experiment (Kelling, 1974a, 1974b) was an example of a major evaluation that relied on an experimental design that was intended to resolve a key policy question: whether the level of routine preventive patrol (assigned randomly to samples of police patrol beats in Kansas City, Missouri) made any differences to the actual and perceived levels of crime and safety in the selected patrol districts of Kansas City. Because the patrol levels were kept “secret” to conceal them from the citizens (and presumably potential law breakers), the experimental results encountered a substantial external validity problem—even if the findings supported the hypothesis that the level of routine preventive patrol had no significant impacts on perceived levels of safety and crime, or on actual levels of crime, how could any other police department announce that it was going to reduce preventive patrol without jeopardizing citizen (and politicians’) confidence? Even in the experiment itself, there was evidence that the police officers who responded to calls for service in the reduced patrol beats did so with more visibility (lights and sirens) than normal—suggesting that they wanted to establish their visibility in the low-patrol beats (Kelling, 1974a; Larson, 1982). In other words, there was
a construct validity threat that was not adequately controlled—compensatory rivalry was operating, whereby the patrol officers in the low-patrol beats acted to “beef up” the perceived level of patrol in their beats.

The Importance of Causality: The Core of the Evaluation Enterprise

Picciotto (2011) points to the centrality of program effectiveness as a core issue for evaluation as a discipline/profession:

What distinguishes evaluation from neighboring disciplines is its unique role in bridging social science theory and policy practice. By focusing on whether a policy, a programme or project is working or not (and unearthing the reasons why by attributing outcomes) evaluation acts as a transmission belt between the academy and the policy-making. (p. 175)

Advocates for experimental research designs point out that since most evaluations include, as a core issue, whether the program was effective, experimental designs are the best and least ambiguous way to answer these causal questions.

Michael Scriven (2008) has taken an active role, since the changes in U.S. policies have privileged RCTs, in challenging the primacy of experimental designs as the methodological backbone of evaluations; in the introduction to a paper published in 2008, he asserts, “The causal wars are still raging, and the amount of collateral damage is increasing” (p. 11). In a series of publications (Cook, Scriven, Coryn, & Evergreen, 2010; Scriven, 2008), he has argued that it is possible to generate valid causal knowledge in many other ways, and argues for a pluralism of methods that are situationally appropriate in the evaluation of programs. In one of his presentations, a key point he makes is that human beings are “hardwired” to look for causal relationships in the world around them. In an evolutionary sense, we have a built-in capacity to observe causal connections. In Scriven’s (2004) words,

Our experience of the world and our part in it, is not only well understood by us but pervasively, essentially, perpetually a causal experience. A thousand times a day we observe causation, directly and validly, accurately and sometimes precisely. We see people riding bicycles, driving trucks, carrying boxes up stairs, turning the pages of books, picking goods off shelves, calling names, and so on. So, the basic kind of causal data, vast quantities of highly reliable and checkable causal data, comes from observation, not from elaborate experiments. Experiments, especially RCTs, are a marvelously ingenious extension of our observational skills, enabling us to infer to causal conclusions where observation alone cannot take us. But it is surely to reverse reality to suppose that they are the primary or only source of reliable causal claims: they are, rather, the realm of flight for such claims, where the causal claims of our everyday lives are the ground traffic of them. (pp. 6–7)

Alternative Perspectives on the Evaluation Enterprise

At the other end of the spectrum are approaches that eschew positivistic or post-positivistic approaches to evaluation and advocate methodologies that are rooted in
anthropology or subfields of sociology or other disciplines. Advocates of these interpretivist (generally qualitative) approaches have pointed out that the positivist view of evaluation is itself based on a set of beliefs about observing and measuring patterns of human interactions. We introduced different approaches to qualitative evaluation in Chapter 5 and pointed out that in the 1980s a different version of “paradigm wars” happened in the evaluation field—that “war” was between the so-called quals and the quants.

Quantitative methods cannot claim to eliminate the need for evaluators to use professional judgment. Smith (1994) argues that quantitative methods necessarily involve judgment calls:

Decisions about what to examine, which questions to explore, which indicators to choose, which participants and stakeholders to tap, how to respond to unanticipated problems in the field, which contrary data to report, and what to do with marginally significant statistical results are judgment calls. As such they are value-laden and hence subjective. . . . Overall the degree of bias that one can control through random assignment or blinded assessment is a minute speck in the cosmos of bias. (pp. 38–39)

Moreover, advocates of qualitative approaches argue that quantitative methods miss the meaning of much of human behavior. Understanding intentions is critical to getting at what the “data” really mean, and the only way to do that is to embrace methods that treat individuals as unique sense-makers who need to be understood on their own terms (Schwandt, 2000).

Kundin (2010) advocates the use of qualitative, naturalistic approaches to understand how evaluators use their “knowledge, experience and judgment to make decisions in their everyday work” (p. 350). Her view is that the methodology-focused logic of inquiry that is often embedded in textbooks does not reflect actual practice:

Although this logic is widely discussed in the evaluation literature, some believe it is rarely relied upon in practice. Instead researchers . . . suggest that evaluators use their intuition, judgment, and experience to understand the evaluand, and by doing so, they understand its merits through an integrated act of perceiving and valuing. (p. 352)

More recently, evaluators who have focused on getting evaluations used have tended to take a pragmatic stance in their approaches, mixing qualitative and quantitative methodologies in ways that are intended to be situationally appropriate (Patton, 2008). They recognize the value of being able to use structured designs where they are feasible and appropriate, but also recognize the value of employing a wide range of complementary (mixed methods) approaches in a given situation to create information that is credible, and hence more likely to be used. We discussed mixed-methods designs in Chapter 5.

Reconciling Evaluation Theory With the Diversity of Practice

The practice of program evaluation is even more diverse than the range of normative approaches and perspectives that populate the textbook and coursework landscape. Experimental evaluations continue to be done and are still viewed by many practitioners as exemplars (Chen, Donaldson, & Mark, 2011; Henry & Mark, 2003). Substantial investments in
time and resources are typically required, and this limits the number and scope of evaluations that are able to randomly assign units of analysis (usually people) to program and control conditions.

Conducting experimental evaluations entails creating a structure that may produce statistical conclusions and internal validity yet fails to inform decisions about implementing the program or policy in non-experimental settings (the Kansas City Preventive Patrol Experiment is an example of that) (Kelling, 1974a; Larson, 1982). Typically, experiments are time limited, and as a result participants can adjust their behaviors to their expectations of how long the experiment will last, as well as to what their incentives are as it is occurring. Cronbach (1982) was eloquent in his criticisms of the (then) emphasis on internal validity as the central criterion for judging the quality of research designs for evaluations of policies and programs. He argued that the uniqueness of experimental settings undermines the extent to which well-constructed (internally valid) experiments can be generalized to other units of analysis, treatments, observing operations, and settings (UTOS). Cronbach argued for the primacy of external validity of evaluations to make them more relevant to policy settings. Shadish, Cook, and Campbell's (2002) book can be seen in part as an effort to address Cronbach’s criticisms of the original Cook and Campbell (1979) book on experimental and quasi-experimental research designs.

The existence of controversies over the construction, execution, and interpretation of many of the large-scale social experiments that were conducted in the 1970s to evaluate programs and policies suggest that very few methodologies are unassailable—even experimental research designs (Basilevsky & Hum, 1984). The craft of evaluation research, even research that is based on randomly assigned treatment and control conditions, is such that its practitioners do not agree about what exemplary practice is, even in a given situation.

In the practice of evaluation, it is rare to have the resources and the control over the program setting needed to conduct even a quasi-experimental evaluation. Instead, typical evaluation settings are limited by significant resource constraints and the expectation that the evaluation process will somehow fit into the existing administrative process that has implemented a policy or a program. The widespread interest in performance measurement as an evaluation approach tends to be associated with an assumption that existing managerial and information technological resources will be sufficient to implement performance measurement systems, and produce information for formative and summative evaluative purposes. We have pointed out the limitations of substituting performance measurement for program evaluation in Chapters 1 and 8 of this textbook.

Working in the Swamp: The Real World of Evaluation Practice

Typical program evaluation methodologies rely on multiple, independent data sources to “strengthen” research designs that are case study or implicit designs (diagrammed in Chapter 3 as XO designs, where X is the program and O is the set of observations on the outcomes that are expected to be affected by the program). The program has been implemented at some time in the past, and now, the evaluator is expected to assess program effectiveness. There is no pretest and no control group; there are insufficient resources to construct these comparisons, and in most situations, comparison groups would not exist. Although multiple data sources permit triangulation of findings, that does not change the
fact that the basic research design is the same; it is simply repeated for each data source (which is a strength since measurement errors would likely be independent) but is still subject to all the weaknesses of that design. In sum, typical program evaluations are conducted after the program is implemented, in settings where the evaluation team has to rely on evidence about the program group alone (i.e., there is no control group). In most evaluation settings, these designs rely on mixed qualitative and quantitative lines of evidence.

In such situations, some evaluators would advocate not using the evaluation results to make any causal inferences about the program. In other words, it is argued that such evaluations ought not to be used to try to address the question: “Did the program make a difference, and if so, what difference(s) did it make?” Instead the evaluation should be limited to addressing the question of whether intended outcomes were actually achieved, regardless of whether the program “produced” those outcomes. That is essentially what performance measurement systems do.

But, many evaluations are commissioned with the need to know whether the program worked and why. Even formative evaluations often include questions about the effectiveness of the program (Cronbach, 1980; Weiss, 1998). Answering the “why” question entails looking at causes and effects.

In situations where a client wants to know if and why the program was effective, and there is clearly insufficient time, money, and control to construct an evaluation design that meets criteria that are textbook-appropriate for answering those questions using an experimental design, evaluators have a choice. They can advise their client that wanting to know whether the program or policy worked—and why—is perhaps not feasible, or they can proceed, understanding that their work may not be as defensible as some textbooks would advocate.

Usually, some variation of the work proceeds. Although comparisons between program and no-program groups are not possible, comparisons among program recipients, comparisons over time for program recipients who have participated in the program, and comparisons among the perspectives of other stakeholders are all possible. We maintain that the way to answer causal questions without research designs that can categorically rule out rival hypotheses is to acknowledge that in addressing issues like program effectiveness (which we take to be the central question in most evaluations), we cannot offer definitive findings or conclusions. Instead, our findings, conclusions, and our recommendations, supported by the evidence at hand and by our professional judgment, will reduce the uncertainty associated with the question.

In this textbook we have argued that in all evaluations, regardless of how sophisticated they are, evaluators use one form or another of professional judgment. The difference between the most sophisticated experimentally designed evaluation and an evaluation based on a case study/implicit design is the amount and the kinds of professional judgments that are entailed—not that the former is appropriate for assessing program effectiveness and the latter is not. Unlike some who have commented on the role of professional judgment in program evaluations and see making judgments as a particular phase in the evaluation process (Skolits, Morrow, & Burr, 2009), we see professional judgment being exercised throughout the entire evaluation process.

Where a research design is (necessarily) weak, we introduce to a greater extent our own experience and our own assessments, which in turn are conditioned by our values, beliefs, and expectations. These become part of the basis on which we interpret the evidence at hand and are also a part of the conclusions and the recommendations. This
professional judgment component in every evaluation means that we should be aware of what it is, and learn how to cultivate sound professional judgment.

Common Ground Between Program Evaluators and Program Managers

The view that all evaluations incorporate professional judgments to a greater or lesser extent means that evaluators have a lot in common with program managers. Managers often conduct assessments of the consequences of their decisions—informal evaluations, if you will. These are not usually based on a necessarily systematic gathering of information, but instead often rely on a manager’s own observations, values, beliefs, expectations, and experiences—their professional judgment. That these assessments are done “on the fly” and are often based on information that is gathered using research designs that do not warrant causal conclusions does not vitiate their being the basis for good management practice. Good managers become skilled at being able to recognize patterns in the complexity of their environments. Inferences from observed or sensed patterns (Mark, Henry, & Julnes, 2000) to causal linkages are informed by their experience and judgment, are tested by observing and often participating in the consequences of a decision, and in turn add to the fund of knowledge and experience that contributes to their capacity to make sound professional judgments.

Situating Professional Judgment in Program Evaluation Practice

Scriven (1994) emphasizes the centrality of judgment (judgments of merit and worth) in the synthesis of evaluation findings/lines of evidence to render an overall assessment of a program. For Scriven, the process of building toward and then rendering a holistic evaluation judgment is a central task for evaluators, a view reflected by others (Skolits et al., 2009). Judgments can be improved by constructing rules or decision processes that make explicit how evidence will be assessed and weighted. Scriven (1994) suggests that, generally, judgments supported by decision criteria are superior to those that are intuitive.

Though evaluations typically use several different lines of evidence to assess a program’s effectiveness and, in the process, have different measures of effectiveness, methodologies such as cost–utility analysis exist for weighting and amalgamating findings that combine multiple measures of program effectiveness (Levin & McEwan, 2001). However, they are data intensive, and apart from the health sector, they are not widely used. The more typical situation is described by House and Howe (1999). They point out that the findings from various lines of evidence in an evaluation may well contain conflicting information about the worth of a program. In this situation, evaluators use their professional judgment to produce an overall conclusion. The process of rendering such a judgment engages the evaluator’s own knowledge, values, beliefs, and expectations. House and Howe (1999) describe this process:

The evaluator is able to take relevant multiple criteria and interests and combine them into all-things-considered judgments in which everything is consolidated and related. . . . Like a referee in a ball game, the evaluator must follow certain sets of rules,
procedures, and considerations—not just anything goes. Although judgment is involved, it is judgment exercised within the constraints of the setting and accepted practice. Two different evaluators might make different determinations, as might two referees, but acceptable interpretations are limited. In the sense that there is room for the evaluator to employ judgment, the deliberative process is individual. In the sense that the situation is constrained, the judgment is professional. (p. 29)

There are also many situations where evaluators must make judgments in the absence of clear methodological constraints or rules to follow. House and Howe (1999) go on to point out that

for evaluators, personal responsibility is a cost of doing business, just as it is for physicians, who must make dozens of clinical judgments each day and hope for the best. The rules and procedures of no profession are explicit enough to prevent this. (p. 30)

Although House and Howe point out that evaluators must make judgments, the process by which judgments are made is nevertheless not well understood. Hurteau, Houle, and Mongiat (2009), in a meta-analysis of 50 evaluation studies, examined the ways that judgments are evidenced and found that in only 20 of those studies had the evaluator(s) made a judgment based on the findings. In addition, in none of those 20 studies do the evaluators describe the process that they used to render the judgment(s).

Program evaluators are currently engaged in debates around the issue of professionalizing evaluation. One element of that debate is whether and how our knowledge and our practice can be codified so that evaluation is viewed as a coherent body of knowledge and skills, and practitioners are seen to be having a consistent set of competencies (King, Stevahn, Ghere, & Minnema, 2001; Stevahn, King, Ghere, & Minnema, 2005b). This debate has focused in part on what is needed to be an effective evaluator—core competencies that provide a framework for assessing the adequacy of evaluation training as well as the adequacy of evaluator practice.

In a study of 31 evaluation professionals in the United States, practitioners were asked to rate the importance of 49 evaluator competencies (King et al., 2001) and then try to come to a consensus about the ratings, given feedback on how their peers had rated each item. The 49 items were grouped into four broad clusters of competencies: (1) systematic inquiry (most items were about methodological knowledge and skills), (2) competent evaluation practice (most items focused on organizational and project management skills), (3) general skills for evaluation practice (most items were on communication, teamwork, and negotiation skills), and (4) evaluation professionalism (most items focused on self-development and training, ethics and standards, and involvement in the evaluation profession).

Among the 49 competencies, one was “making judgments” and referred to an overall evaluative judgment, as opposed to recommendations, at the end of an evaluation (King et al., 2001, p. 233). It was rated the second lowest on average among all the competencies. This finding suggests that judgment, comparatively, is not perceived to be that important (although the item average was still 74.68 out of 100 possible points). King et al. (2001) suggested that “some evaluators agreed with Michael Scriven that to evaluate is to judge; others did not” (p. 245). The “reflects on practice” item, however, was given an average rating of 93.23—a ranking of 17 among the 49 items. Schön (1987) makes reflection on
one’s practice the key element in being able to develop sound professional judgment. For both of these items, there was substantial variation among the practitioners about their ratings, with individual ratings ranging from 100 (highest possible score) to 20. The discrepancy between the low overall score for “making judgments” and the higher score for “reflects on practice” may be related to the difference between making a judgment, as an action, and reflecting on practice, as a personal quality.

We see professional judgment being a part of the whole process of working with clients, framing evaluation questions, designing and conducting evaluation research, analyzing and interpreting the information, and communicating the findings, conclusions, and recommendations to stakeholders. If you go back to the outline of a program evaluation process offered in Chapter 1, or the outline of the design and implementation of a performance measurement system offered in Chapter 9, you will see professional judgment is a part of all the steps in both processes. Furthermore, we see different kinds of professional judgment being more or less important at different stages in evaluation processes. We will come back to the relationships between evaluation competencies and professional judgment later in this chapter.

ACQUIRING KNOWLEDGE AND SKILLS FOR EVALUATION PRACTICE

The idea that evaluation is a profession, or aspires to be a profession, is an important part of contemporary discussions of the scope and direction of the enterprise (Altschuld, 1999). Modarresi, Newman, and Abolafia (2001) quote Leonard Bickman (1997), who was president of the American Evaluation Association (AEA) in 1997, in asserting that “we need to move ahead with professionalizing evaluation or else we will just drift into oblivion” (p. 1). Bickman and others in the evaluation field were aware that other related professions continue to carve out territory, sometimes at the expense of evaluators. Picciotto (2011) points out, however, that “heated doctrinal disputes within the membership of the AEA have blocked progress [towards professionalization] in the USA” (p. 165).

What does it mean to be a professional? What distinguishes a profession from other occupations? Eraut (1994) suggests that professions are characterized by the following: a core body of knowledge that is shared through the training and education of those in the profession; some kind of government-sanctioned license to practice; a code of ethics and standards of practice; and self-regulation (and sanctions for wrongdoings) through some kind of professional association to which members of the practice community must belong.

Professional Knowledge as Applied Theory

The core body of knowledge that is shared among members of a profession can be characterized as knowledge that is codified, publicly available (taught for and learned by aspiring members of the profession), and supported by validated theory (Eraut, 1994). One view of professional practice is that competent members of a profession apply this validated theoretical knowledge in their work. Competent practitioners are persons who have the
credentials of the profession (including evidence that they have requisite knowledge) and have established a reputation for being able to translate theoretical knowledge into sound practice.

Professional Knowledge as Practical Know-How

An alternative view of professional knowledge is that it is the application of practical know-how to particular situations. The competent practitioner uses his or her experiential and intuitive knowledge to assess a situation and offer a diagnosis (in the health field) or a decision in other professions (Eraut, 1994). Although theoretical knowledge is a part of what competent practitioners rely on in their work, practice is seen as more than applying theoretical knowledge. It includes a substantial component that is learned through practice itself. Although some of this knowledge can be codified and shared (Schön, 1987; Tripp, 1993), part of it is tacit, that is, known to individual practitioners, but not shareable in the same ways that we share the knowledge in textbooks, lectures, or other publicly accessible learning and teaching modalities (Schwandt, 2008).

Polanyi (1958) described tacit knowledge as the capacity we have as human beings to integrate “facts” (data and perceptions) into patterns. He defined tacit knowledge in terms of the process of discovering theory: “This act of integration, which we can identify both in the visual perception of objects and in the discovery of scientific theories, is the tacit power we have been looking for. I shall call it tacit knowing” (Polanyi & Grene, 1969, p. 140).

For Polanyi, tacit knowledge cannot be communicated directly. It has to be learned through one’s own experiences—it is by definition personal knowledge. Knowing how to ride a bicycle, for example, is in part tacit. We can describe to others how the physics and the mechanics of getting onto a bicycle and riding it works, but the experience of getting onto the bicycle, pedaling, and getting it to stay up is quite different from being told how to do so.

One implication of acknowledging that what we know is in part personal is that we cannot teach everything that is needed to learn a skill. The learner can be guided with textbooks, good examples, and even demonstrations, but that knowledge (Polanyi calls it impersonal knowledge) must be combined with the learner’s own capacity to tacitly know—to experience the realization (or a series of them) that he or she understands how to use the skill.

Clearly, from this point of view, practice is an essential part of learning. One’s own experience is essential for fully integrating impersonal knowledge into working knowledge. But because the skill that has been learned is in part tacit, when the learner tries to communicate it, he or she will discover that, at some point, the best advice is to suggest that the new learner try it and “learn by doing.” This is a key part of craftsmanship.

Balancing Theoretical and Practical Knowledge in Professional Practice

The difference between the applied theory and the practical know-how views of professional knowledge (Fish & Coles, 1998; Schwandt, 2008) has been characterized as the difference between knowing that (publicly accessible, propositional knowledge and skills)
and knowing how (practical, intuitive, experientially grounded knowledge that involves wisdom, or what Aristotle called praxis) (Eraut, 1994).

These two views of professional knowledge highlight different views of what professional practice is and indeed ought to be. The first view can be illustrated with an example. In the field of medicine, the technical/rational view of professional knowledge and professional practice continues to support efforts to construct and use expert systems—software systems that can offer a diagnosis based on a logic model that links combinations of symptoms in a probabilistic tree to possible diagnoses (Fish & Coles, 1998). By inputting the symptoms that are either observed or reported by the patient, the expert system (embodying the public knowledge that is presumably available to competent practitioners) can treat the diagnosis as a problem to solve. Clinical decision making employs algorithms that produce a probabilistic assessment of the likelihood that symptom, drug, and other technical information will support one or another alternative diagnoses.

The second view of professional knowledge as practical know-how embraces a view of professional practice as craftsmanship and artistry. Although it acknowledges the importance of experience in becoming a competent practitioner, it also complicates our efforts to understand the nature of professional practice. If practitioners know things that they cannot share and their knowledge is an essential part of sound practice, how do professions find ways of ensuring that their members are competent?

Schwandt (2008) recognizes the importance of balancing applied theory and practical knowledge in evaluation. His concern is with the tendency, particularly in performance management systems where practice is circumscribed by a focus on outputs and outcomes, to force “good practice” to conform to some set of performance measures and performance results:

The fundamental distinction between instrumental reason as the hallmark of technical knowledge and judgment as the defining characteristic of practical knowledge is instinctively recognizable to many practitioners . . . Yet the idea that “good” practice depends in a significant way on the experiential, existential, knowledge we speak of as perceptivity, insightfulness, and deliberative judgment is always in danger of being overrun by (or at least regarded as inferior to) an ideal of “good” practice grounded in notions of objectivity, control, predictability, generalizability beyond specific circumstances, and unambiguous criteria for establishing accountability and success. This danger seems to be particularly acute of late, as notions of auditable performance, output measurement, and quality assurance have come to dominate the ways in which human services are defined and evaluated. (p. 37)

The idea of balance is further explored in the section below, where we discuss various aspects of professional judgment.

**UNDERSTANDING PROFESSIONAL JUDGMENT**

What are the different kinds of professional judgment? How does professional judgment impact the range of decisions that evaluators make? Can we construct a model of how professional judgment affects evaluation-related decisions?
Fish and Coles (1998) have constructed a typology of four kinds of professional judgment in the health care field. We believe that these can be generalized to the evaluation field. Each builds on the previous one; the extent and kinds of judgment differ across the four kinds. At one end of the continuum, practitioners apply technical judgments that are about specific issues involving routine tasks. Typical questions include the following: What do I do now? How do I apply my existing knowledge and skills to do this routine task? In an evaluation, an example of this kind of judgment would be how to select a random sample from a population of case files in a social service agency.

The next level is procedural judgment, which focuses on procedural questions and involves the practitioner comparing the skills/tools that he or she has available to accomplish a task. Practitioners ask questions like “What are my choices to do this task?” “From among the tools/knowledge/skills available to me, which combination works best for this task?” An example from an evaluation would be deciding how to contact clients in a social service agency—whether to use a survey (and if so, mailing, telephone, interview format, or some combination) or use focus groups (and if so, how many, where, how many participants in each, how to gather them).

The third level of professional judgment is reflective. It again assumes that the task or the problem is a given, but now the practitioner is asking the following questions: How do I tackle this task? Given what I know, what are the ways that I could proceed? Are the tools that are easily within reach adequate, or instead, should I be trying some new combination or perhaps developing some new ways of dealing with this task or problem? A defining characteristic of this third level of professional judgment is that the practitioner is reflecting on his or her practice and seeking ways to enhance his or her practical knowledge and skills and perhaps innovate to address a given situation.

An example from a needs assessment for child sexual abuse prevention programs in an urban school district serves to illustrate reflective judgment on the part of the evaluator in deciding on the research methodology. Classes from an elementary school are invited to attend a play acted by school children of the same ages as the audience. The play is called “No More Secrets” and is about an adult–child relationship that involves touching and other activities. At one point in the play, the “adult” tells the “child” that their touching games will be their secret. The play is introduced by a professional counselor, and after the play, children are invited to write questions on cards that the counselor will answer. The children are told that if they have questions about their own relationships with adults, these questions will be answered confidentially by the counselor. The evaluator, having obtained written permissions from the parents, contacts the counselor, who, without revealing the identities of any of the children, indicates to the evaluator the number of potentially abusive situations among the students who attended the play. Knowing the proportion of the school district students that attended the play, the evaluator is able to roughly estimate the incidence of potentially abusive situations in that school-age population.

The fourth level of professional judgment is deliberative—it explicitly involves a practitioner’s own values. Here the practitioner is asking the following question: What ought I to be doing in this situation? No longer are the ends or the tasks fixed, but instead the professional is taking a broad view that includes the possibility that the task or problem may or may not be an appropriate one to pursue. Professionals at this level are asking questions about the nature of their practice and connecting what they do as professionals
with their broader values and moral standards. We discuss evaluation ethics later in this chapter. The case study in Appendix A of this chapter is an example of a situation that would involve deliberative judgment.

A Modeling of the Professional Judgment Process

Since professional judgment spans the evaluation process, it will influence a wide range of decisions that evaluators make in their practice. The four types of professional judgment that Fish and Coles (1998) describe suggest decisions of increasing complexity from discrete technical decisions to global decisions that can affect an evaluator’s present and future roles as an evaluation practitioner. Figure 12.1 displays a model of the way that professional judgment is involved in evaluator decision making. The model focuses on a single decision—a typical evaluation would involve many such decisions of varying complexity. In the model, evaluator values, beliefs, and expectations, together with both shareable and practical (tacit) knowledge combine to create a fund of experience that is tapped for professional judgments. In turn, professional judgments influence the decision at hand.

We will present the model and then discuss it, elaborating on the meanings of the key constructs in the model.

Evaluator decisions have consequences. They may be small—choosing a particular alpha (α) level for tests of statistical significance will have an impact on which findings are noteworthy, given a criterion that significant findings are worth reporting; or they may be large—deciding not to conduct an evaluation in a situation where the desired findings are being specified in advance by a key stakeholder could affect the career of an evaluation practitioner. These consequences feed back to our knowledge (both our shareable and our practical know-how), values, beliefs, and expectations. Evaluators have an opportunity to learn from each decision, and one of our challenges as professionals is to increase

![Figure 12.1](image-url)
the likelihood that we take advantage of such learning opportunities. We will discuss
reflective practice later in this chapter.

The model can be unpacked by discussing key constructs in it. Some constructs have
been elaborated in this chapter already (shareable knowledge, practical know-how, and
professional judgment), but it is worthwhile to define each one explicitly in one table. Table 12.1
summarizes the constructs in Figure 12.1 and offers a short definition of each. Several of the
constructs will then be discussed further to help us understand what roles they play in the
process of forming and applying professional judgment.

<table>
<thead>
<tr>
<th>Constructs in the Model</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>Values are statements about what is desirable, what ought to be, in a given situation.</td>
</tr>
<tr>
<td>Beliefs</td>
<td>Beliefs are about what we take to be true—our assumptions about how we know what we know (our epistemologies are examples of beliefs).</td>
</tr>
<tr>
<td>Expectations</td>
<td>Expectations are assumptions that are typically based on what we have learned and what we have come to accept as normal. Expectations can limit what we are able to “see” in particular situations.</td>
</tr>
<tr>
<td>Shareable knowledge</td>
<td>Knowledge that is typically found in textbooks or other such media; knowledge that forms the core of the formal training and education of professionals in a field.</td>
</tr>
<tr>
<td>Practical know-how</td>
<td>Practical know-how is the knowledge that is gained through practice. It complements shareable knowledge and can be tacit—that is, acquired from one’s professional practice and not shareable.</td>
</tr>
<tr>
<td>Experience</td>
<td>Experience is an amalgam of our knowledge, values, beliefs, expectations, and practical know-how. For a given decision, we have a “fund” of experience that we can draw from. We can augment that fund with learning, and from the consequences of the decisions we make as professionals.</td>
</tr>
<tr>
<td>Professional judgment</td>
<td>Professional judgment is a process that relies on our experience and ranges from technical judgments to deliberative judgments.</td>
</tr>
<tr>
<td>Decision</td>
<td>In a typical evaluation, evaluators make hundreds of decisions that collectively define the evaluation process. Decisions are choices—a choice made by an evaluator about everything from discrete methodological issues to global values–based decisions that affect the whole evaluation (and perhaps future evaluations).</td>
</tr>
<tr>
<td>Consequences</td>
<td>Each decision has consequences—for the evaluator and for the evaluation process. Consequences can range from discrete to global, commensurate with the scope and implications of the decision.</td>
</tr>
<tr>
<td>Decision environment</td>
<td>The decision environment is the set of factors that influences the decision-making process, including the stock of knowledge that is available to the evaluator. Among the factors that could impact an evaluator decision are professional standards, resources (including time and data), incentives (perceived consequences that induce a particular pattern of behavior), and constraints (legal, institutional, and regulatory requirements that specify the ways that evaluator decisions must fit a decision environment).</td>
</tr>
</tbody>
</table>
The Decision Environment

The particular situation or problem at hand influences how a program evaluator’s professional judgment will be exercised. Each opportunity for professional judgment will have unique characteristics that will demand that it be approached in particular ways. For example, a methodological issue will require a different kind of professional judgment from one that centers on an ethical issue. Even two cases involving a similar question of methodological choice will have facts about each of them that will influence the professional judgment process. We would agree with evaluators who argue that methodologies need to be situationally appropriate, avoiding a one-size-fits-all approach (Patton, 2008). The extent to which the relevant information about a particular situation is known or understood by the evaluator will affect the professional judgment process.

The decision environment includes constraints and incentives and costs and benefits, both real and perceived, that affect professional judgment. Some examples include the expectations of the client, the professional’s lines of accountability, tight deadlines, complex and conflicting objectives, and financial constraints. For people working within an organization—for example, internal evaluators—the organization also presents a significant set of environmental factors, in that its particular culture, goals, and objectives may have an impact on the way the professional judgment process operates.

Relevant professional principles and standards such as the AEA’s (2004) “Guiding Principles for Evaluators” also form part of the judgment environment because, to some extent, they interact with and condition the free exercise of judgment by professionals and replace individual judgment with collective judgment (Gibbins & Mason, 1988). We will come back to evaluation standards later in this chapter.

Values, Beliefs, and Expectations

Professional judgment is influenced by personal characteristics of the person exercising it. It must always be kept in mind that “judgment is a human process, with logical, psychological, social, legal, and even political overtones” (Gibbins & Mason, 1988, p. 18). Each of us has a unique combination of values, beliefs, and expectations that make us who we are, and each of us has internalized a set of professional norms that make us the kind of practitioner that we are. These personal factors can lead two professionals to make quite different professional judgments about the same situation (Tripp, 1993).

Among the personal characteristics that can influence one’s professional judgment, expectations are among the most important. Expectations have been linked to paradigms; perceptual and theoretical structures that function as frameworks for organizing one’s perspectives, even one’s beliefs about what is real and what is taken to be factual. Kuhn (1962) has suggested that paradigms are formed through our education and training. Eraut (1994) has suggested that the process of learning to become a professional is akin to absorbing an ideology.

Our past experiences (including the consequences of previous decisions we have made in our practice) predispose us to understand or even expect some things and not others, to interpret situations, and consequently to behave in certain ways rather than in others. As Abercrombie (1960) argues, “We never come to an act of perception with an entirely blank mind, but are always in a state of preparedness or expectancy, because of our past experiences” (p. 53). Thus, when we are confronted with a new situation, we perceive and interpret it in whatever way makes it most consistent with our existing understanding of the world, with
our existing paradigms. For the most part, we perform this act unconsciously. We are not aware of how our particular worldview influences how we interpret and judge the information we receive on a daily basis in the course of our work or how it affects our subsequent behavior.

How does this relate to our professional judgment? Our expectations lead us to see things we are expecting to see, even if they are not actually there, and to not see things we are not expecting, even if they are there. Abercrombie (1960) calls our worldview our “schemata” and illustrates its power over our judgment process with the following figure (Figure 12.2).

In most cases, when we first read the phrases contained in the triangles, we do not see the extra words. As Abercrombie (1960) points out, “it’s as though the phrase ‘Paris in the Spring,’ if seen often enough, leaves a kind of imprint on the mind’s eye, into which the phrase in the triangle must be made to fit” (p. 35). She argues that “if [one’s] schemata are not sufficiently ‘living and flexible,’ they hinder instead of help [one] to see” (p. 29). Our tendency is to ignore or reject what does not fit our expectations. Thus, similar to the way we assume the phrases in the triangles make sense and therefore unconsciously ignore the extra words, our professional judgments are based in part on our preconceptions and thus may not be appropriate for the situation.

Expectations can also contribute to improving our judgment by allowing us to unconsciously know how best to act in a situation. When the consequences of such a decision are judged to be salutary, our expectations are reinforced.

Acquiring Professional Knowledge

Our professional training and education are key influences; they can affect professional judgment in positive ways by not only allowing us to understand and address problems in a manner that those without the same education could not, but they also predispose us to interpret situations in particular ways. Indeed, professional education is often one of the most pervasive reasons for our acceptance of “tried and true” ways of approaching problems in professional practice. As Katz (1988) observes, “Conformity and orthodoxy, playing the game according to the tenets of the group to which students wish to belong, are encouraged in . . . all professional education” (p. 552). Thus, somewhat contrary to what would appear to be common sense, professional judgment does not
necessarily improve in proportion to increases in professional training and education. Similarly, professional judgment does not necessarily improve with increased professional experience, if such experience does not challenge but only reinforces already accepted ideologies. Ayton (1998) makes the point that even experts in a profession are not immune to poor professional judgment:

One view of human judgment is that people—including experts—not only suffer various forms of myopia but are somewhat oblivious of the fact. . . . Experts appear to have very little insight into their own judgment. . . . This oblivion in turn might plausibly be responsible for further problems, e.g. overconfidence . . . attributed, at least in part, to a failure to recognize the fallibility of our own judgment. (pp. 238–239)

On the other hand, Mowen (1993) notes that our experience, if used reflectively and analytically to inform our decisions, can be an extremely positive factor contributing to good professional judgment. Indeed, he goes so far as to argue that “one cannot become a peerless decision maker without that well-worn coat of experience . . . the bumps and bruises received from making decisions and seeing their outcomes, both good or bad, are the hallmark of peerless decision makers” (p. 243).

**IMPROVING PROFESSIONAL JUDGMENT IN EVALUATION THROUGH REFLECTIVE PRACTICE**

Having reviewed the ways that professional judgment is woven through the fabric of evaluation practice and having shown how professional judgment plays a part in our decisions as evaluation practitioners, we can turn to discussing ways of self-consciously improving our professional judgment. Key to this process is becoming aware of one’s own decision-making processes.

**Guidelines for the Practitioner**

Epstein (1999) suggests that a useful stance for professional practice is mindfulness. Krasner et al. (2009) define mindfulness this way:

The term *mindfulness* refers to a quality of awareness that includes the ability to pay attention in a particular way: on purpose, in the present moment, and nonjudgmentally. Mindfulness includes the capacity for lowering one’s own reactivity to challenging experiences; the ability to notice, observe, and experience bodily sensations, thoughts, and feelings even though they may be unpleasant; acting with awareness and attention (not being on autopilot); and focusing on experience, not on the labels or judgments applied to them. (p. 1285)

Epstein and others have developed programs to help medical practitioners become more mindful (Krasner et al., 2009). In a study involving 70 primary care practitioners in Rochester, New York, participants were trained through an 8-week combination of
weekly sessions and an all-day session to become more self-aware. The training was accompanied by opportunities to write brief stories to reflect on their practice and to use **appreciative inquiry** to identify ways that they had been successful in working through challenging practice situations.

The before-versus-after results suggested that for the doctors “increases in mindfulness correlated with reductions in burnout and total mood disturbance. The intervention was also associated with increased trait emotional stability (i.e. greater resilience)” (p. 1290).

Mindfulness is the cultivation of a capacity to observe, in a nonjudgmental way, one’s own physical and mental processes during and after tasks. In other words, it is the capacity for self-reflection that facilitates bringing to consciousness our values, assumptions, expectations, beliefs, and even what is tacit in our practice. Epstein (1999) suggests, “Mindfulness informs all types of professionally relevant knowledge, including propositional facts, personal experiences, processes, and know-how each of which may be tacit or explicit” (p. 833).

Although mindfulness can be linked to religious and philosophical traditions, it is a secular way of approaching professional practice that offers opportunities to continue to learn and improve (Epstein, 2003). A mindful practitioner is one who has cultivated the art of self-observation (cultivating the compassionate observer). Epstein characterizes mindful practice this way:

> When practicing mindfully, clinicians approach their everyday tasks with critical curiosity. They are present in the moment, seemingly undistracted, able to listen before expressing an opinion, and able to be calm even if they are doing several things at once. These qualities are considered by many to be prerequisite for compassionate care. (p. 2)

The objective of mindfulness is to see what is, rather than what one wants to see or even expects to see. Mindful self-monitoring involves several things: “access to internal and external data; lowered reactivity to inner experiences such as thoughts and emotions; active and attentive observation of sensations, images, feelings, and thoughts; curiosity; adopting a nonjudgmental stance; presence, [that is] acting with awareness . . . ; openness to possibility; adopting more than one perspective; [and] ability to describe one’s inner experience” (Epstein, Siegel, & Silberman, 2008, p. 10).

Epstein (1999) suggests that there are at least three ways of nurturing mindfulness: (1) mentorships with practitioners who are themselves well regarded in the profession; (2) reviewing one’s own work, taking a nonjudgmental stance; and (3) meditation to cultivate a capacity to observe one’s self.

In order to cultivate the capacity to make sound professional judgments it is essential to become aware of the unconscious values and other personal factors that may be influencing one’s professional judgment. For only through coming to realize how much our professional judgments are influenced by these personal factors can we become more self-aware and work toward extending our conscious control of them and their impacts on our judgment. As Tripp (1993) argues, “Without knowing who we are and why we do things, we cannot develop professionally” (p. 54). By increasing our understanding of the way we make professional judgments, we improve our ability to reach deliberate, fully thought-out decisions rather than simply accepting as correct the first conclusion that intuitively comes to mind.
But how can we, as individuals, learn what factors are influencing our own personal professional judgment? One way is to conduct a systematic questioning of professional practice (Fish & Coles, 1998). Professionals should consistently reflect on what they have done in the course of their work and then investigate the issues that arise from this review. Reflection should involve articulating and defining the underlying principles and rationale behind our professional actions and should focus on discovering the “intuitive knowing implicit in the action” (Schön, 1988, p. 69).

Tripp (1993) suggests that this process of reflection can be accomplished by selecting and then analyzing critical incidents that have occurred during our professional practice in the past (critical incident analysis). A critical incident can be any incident that occurred in the course of our practice that sticks in our mind and hence, provides an opportunity to learn. What makes it critical is the reflection and analysis that we bring to it. Through the process of critical incident analysis, we can gain an increasingly better understanding of the factors that have influenced our professional judgments. As Fish and Coles (1998) point out,

Any professional practitioner setting out to offer and reflect upon an autobiographical incident from any aspect of professional practice is, we think, likely to come sooner or later to recognize in it the judgments he or she made and be brought to review them. (p. 254)

For it is only in retrospect, in analyzing our past decisions, that we can see the complexities underlying what at the time may have appeared to be a straightforward, intuitive professional judgment. “By uncovering our judgments . . . and reflecting upon them,” Fish and Coles (1998) maintain, “we believe that it is possible to develop our judgments because we understand more about them and about how we as individuals come to them” (p. 285).

Jewiss and Clark-Keefe (2007) connect reflective practice for evaluators to developing cultural competence. The Guiding Principles for Evaluators (AEA, 2004) makes cultural competence, “seeking awareness of their own culturally-based assumptions, their understanding of the worldviews of culturally-different participants and stakeholders in the evaluation,” one of the core competencies for evaluators. They believe that to become more culturally competent, evaluators would benefit from taking a constructivist stance in reflecting on their own practice:

Constructivism has indeed helped signal evaluators’ responsibility for looking out[ward]: for attending to and privileging program participants’ expressions as the lens through which to learn about, change, and represent programs. Just as important, constructivism conveys evaluators’ responsibilities for looking in[ward]: for working to develop and maintain a critically self-reflective stance to examine personal perspectives and to monitor bias. (p. 337)

Self-consciously challenging the routines of our practice, the “high hard ground” that Schön refers to in the quote at the outset of this chapter, is an effective way to begin to develop a more mindful stance. In our professional practice, each of us will have developed routines for addressing situations that occur frequently. As Tripp (1993) points out, although routines

may originally have been consciously planned and practiced, they will have become habitual, and so unconscious, as expertise is gained over time. Indeed, our routines
often become such well-established habits that we often cannot say why we did one thing rather than another, but tend to put it down to some kind of mystery such as “professional intuition.” (p. 17)

Another key way to critically reflect on our professional practice and understand what factors influence the formation of our professional judgments is to discuss our practice with our colleagues. Colleagues, especially those who are removed from the situation at hand or under discussion, can act as “critical friends” and can help in the work of analyzing and critiquing our professional judgments with an eye to improving them. With different education, training, and experience, our professional peers often have different perspectives from us. Consequently, involving colleagues in the process of analyzing and critiquing our professional practice allows us to compare with other professionals our ways of interpreting situations and choosing alternatives for action. Moreover, the simple act of describing and summarizing an issue so that our colleagues can understand it can reveal and provide much insight into the professional judgments we have incorporated.

The Range of Professional Judgment Skills

There is considerable interest in the evaluation field in outlining competencies that define sound practice (Ghere, King, Stevahn, & Minnema, 2006; King et al., 2001; Stevahn, King, Ghere, & Minnema, 2005a). Although there are different versions of what these competencies are, there is little emphasis on acquiring professional judgment skills as a distinct competency. Efforts to establish whether practitioners themselves see judgment skills as being important indicate a broad range of views, reflecting some important differences as to what evaluation practice is and ought to be (King et al., 2001).

If we consider linkages between types of professional judgment and the range of activities that comprise evaluation practice, we can see that some kinds of professional judgment are more important for some clusters of activities than others. But for many evaluation activities, several different kinds of professional judgment can be relevant. Table 12.2 summarizes clusters of activities that reflect the design and implementation of a typical program evaluation or a performance measurement system. These clusters are based on the outlines for the design and implementation of program evaluations and performance measurement systems included in Chapters 1 and 9, respectively. Although they are not comprehensive, that is, do not absolutely represent the detailed range of activities discussed earlier in this textbook, they illustrate the ubiquity of professional judgment in all areas of our practice.

Table 12.2 suggests that for most clusters of evaluation activities, several different types of professional judgment are in play. The notion that somehow we could practice by exercising only technical and procedural professional judgment, or confining our judgment calls to one part of the evaluation process, is akin to staying on Schön’s (1987) “high hard ground.”

Ways of Improving Sound Professional Judgment Through Education and Training-Related Activities

Developing sound professional judgment depends substantially on being able to develop and practice the craft of evaluation. Schön (1987) and Tripp (1993), among
Table 12.2 Types of Professional Judgment That Are Relevant to Program Evaluation and Performance Measurement

<table>
<thead>
<tr>
<th>Clusters of program evaluation knowledge and skills</th>
<th>Types of Professional Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negotiating the terms of reference for the evaluation</td>
<td>Technical</td>
</tr>
<tr>
<td>Assessing relevant previous work</td>
<td>Yes</td>
</tr>
<tr>
<td>Describing the program and its environment</td>
<td>Yes</td>
</tr>
<tr>
<td>Selecting evaluation methods</td>
<td>Yes</td>
</tr>
<tr>
<td>Assessing evaluability options</td>
<td>Yes</td>
</tr>
<tr>
<td>Conducting the evaluation</td>
<td>Yes</td>
</tr>
<tr>
<td>Reporting and disseminating the results</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clusters of performance measurement knowledge and skills</th>
<th>Types of Professional Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessing organizational capacity to design, implement, and sustain performance measurement systems</td>
<td>Yes</td>
</tr>
<tr>
<td>Identifying performance measures and the information to support them</td>
<td>Yes</td>
</tr>
<tr>
<td>Analysis and reporting of performance results</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusting the performance measurement system over time</td>
<td>Yes</td>
</tr>
</tbody>
</table>

others, have emphasized the importance of practice as a way of cultivating sound professional judgment. Although textbook knowledge (“knowing what”) is also an essential part of every evaluator’s toolkit, a key part of evaluation curricula are opportunities to acquire experience.

There are at least six complementary ways that evaluation curricula can be focused to provide opportunities for students to develop their judgment skills. Some activities are more discrete, that is, are relevant for developing skills that are specific. These are generally limited to a single course or even a part of a course. Others are more generic, offering opportunities to acquire experience that spans entire evaluation processes. These are typically activities that integrate coursework in real work experiences. Table 12.3 summarizes ways that academic programs can inculcate professional judgment capacities in their students.

The types of learning activities in Table 12.3 are typical of many programs that train evaluators, but what is important is realizing that each of these kinds of activities contributes
directly to developing a set of skills that all practitioners need and will use in all their professional work. In an important way, identifying these learning activities amounts to making explicit what has largely been tacit in our profession.

### Teamwork and Professional Judgment

Evaluators and managers often work in organizational settings where teamwork is expected. Successful teamwork requires establishing norms and expectations that encourage good communication, sharing of information, and a joint commitment to the task at hand. Being able to select team members and foster a work environment wherein people are willing to trust each other, and be open and honest about their own views on issues, is conducive to generating information that reflects a diversity of perspectives. Even though there will still be individual biases, the views expressed are more likely to be valid than simply the perceptions of a dominant individual or coalition in the group. An organizational culture that emulates features of learning organizations (Garvin, 1993; Mayne, 2008) will tend to produce information that is more valid as input for making decisions and evaluating policies and programs.

Managers and evaluators who have the skills and experience to be able to call on others and, in doing so, be reasonably confident that honest views about an issue are being offered, have a powerful tool to complement their own knowledge and experience and
their systematic inquiries. Good professional judgment, therefore, is partly about selecting and rewarding people who themselves have demonstrated a capacity to deliver sound professional judgment.

Evaluation as a Craft: Implications for Learning to Become an Evaluation Practitioner

Evaluation has both a methodological aspect, where practitioners are applying tools, albeit with the knowledge that the tools may not fit the situation exactly, and an aesthetic aspect, which entails developing an appreciation for the art of design, the conduct of evaluation-related research, and the interpretation of results. As Berk and Rossi (1999) contend, mastering a craft involves more than learning the techniques and tools of the profession; it involves developing “intelligence, experience, perseverance, and a touch of whimsy” (p. 99), which all form part of professional judgment. Traditionally, persons learning a craft apprenticed themselves to more senior members of the trade. They learned by doing, with the guidance and experience of the master craftsperson.

We have come to think that evaluation can be taught in classrooms, often in university settings or in professional development settings. Although these experiences are useful, they are no substitute for learning how evaluations are actually done. Apprenticing to a person or persons who are competent senior practitioners is an important part of becoming a practitioner of the craft. Some evaluators apprentice themselves in graduate programs, preparing master’s or doctoral theses with seasoned practitioners. Others work with practitioners in work experience settings (e.g., co-op placements). Still others join a company or organization at a junior level and, with time and experience, assume the role of full members of the profession.

Apprenticeship complements what can be learned in classrooms, from textbooks and other such sources. Schön (1987) points out that an ideal way to learn a profession is to participate in practical projects wherein students design for actual situations, under the guidance of instructors or coaches who are themselves seasoned practitioners. Students then learn by doing and also have opportunities, with the guidance of coaches, to critically reflect on their practice.

In evaluation, an example of such an opportunity might be a course that is designed as a hands-on workshop to learn how to design and conduct a program evaluation. Cooksy (2008) describes such a course at Portland State University. Students work with an instructor who arranges for client organizations, who want evaluations done, to participate in the course. Students work in teams, and teams are matched with clients. As the course progresses, each team is introduced to the skills that are needed to meet client and instructor expectations for that part of the evaluation process. There are tutorials to learn skills that are needed for the teams’ work, and opportunities for teams to meet as a class to share their experiences and learn from each other and the instructor. Clients are invited into these sessions to participate as stakeholders and provide the class and the instructor with relevant and timely feedback. The teams are expected to gather relevant lines of evidence, once their evaluation is designed, and analyze the evidence. Written reports for the clients are the main deliverables for the teams, together with oral presentations of the key results and recommendations in class, with the clients in attendance.
The Development of Ethics for Evaluation Practice

In this chapter we have alluded to ethical decision-making as a part of the work evaluators do; it is a consideration in how they exercise professional judgment. The evaluation guidelines, standards, and principles that have been developed for the evaluation profession all speak, in different ways, to ethical practice. Although evaluation practice is not guided by a set of professional norms that are enforceable (Rossi, Lipsey, & Freeman, 2004), ethical guidelines are an important reference point for evaluators. Increasingly, organizations that involve people (e.g., clients or employees) in research are expected to take into account the rights of their participants across the stages of the evaluation: As the study objectives are framed, measures and data collection are designed and implemented, results are interpreted, and findings are disseminated. In universities, human research ethics committees routinely scrutinize research plans to ensure that they do not violate the rights of participants. In both the United States and Canada, there are national policies or regulations that are intended to protect the rights of persons who are participants in research (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada, 2010; U.S. Department of Health and Human Services, 2009).

The past quarter century has witnessed major developments in the domain of evaluation ethics. These include publication of the original and revised versions of the Guiding Principles for Evaluators (AEA, 1995, 2004), and the second and third editions of the Program Evaluation Standards (Sanders, 1994; Yarbrough, Shulha, Hopson, & Caruthers, 2011). Two examples of books devoted to program evaluation ethics (Morris, 2008; Newman & Brown, 1996) as well as chapters on ethics in handbooks in the field (Seiber, 2009; Simons, 2006) are additional resources. The AEA is active in promoting evaluation ethics with the creation of the Ethical Challenges section of the American Journal of Evaluation (Morris, 1998) and the addition of an ethics training module to the website of the AEA, as described in Morris’s The Good, the Bad, and the Evaluator: 25 Years of AJE Ethics (Morris, 2011).

Morris (2011) has followed the development of evaluation ethics over the past quarter century and notes that there are few empirical studies that focus on evaluation ethics to date. Additionally, he argues that “most of what we know (or think we know) about evaluation ethics comes from the testimonies and reflections of evaluators”—leaving out the crucial perspectives of other stakeholders in the evaluation process (p. 145). Textbooks on the topic of evaluation range in the amount of attention that is paid to evaluation ethics—in some textbooks, it is the first topic of discussion on which the rest of the chapters rest, as in, for example, Qualitative Researching by Jennifer Mason (2002). In others, the topic arises later, or in some cases it is left out entirely.

Newman and Brown (1996) have undertaken an extensive study of evaluation practice to establish ethical principles that are important for evaluators in the roles they play. Underlying their work are principles, which they trace to Kitchener’s (1984) discussions of ethical norms. Table 12.4 summarizes ethical principles that are taken in part from Newman and Brown (1996) and from the Tri-Council Policy on the Ethical Conduct for Research Involving...
Humans (Canadian Institutes of Health Research et al., 2010), and shows they correspond to the AEA’s Guiding Principles for Evaluators (AEA, 2004) and the Canadian Evaluation Society (CES) Guidelines for Ethical Conduct (CES, 2012a).

The ethical principles summarized in Table 12.4 are not absolute and arguably are not complete. Each one needs to be weighed in the context of a particular evaluation project and balanced with other ethical considerations. For example, the “keeping promises” principle suggests that contracts, once made, are to be honored, and normally that is the case. But consider the following example: An evaluator makes an agreement with the executive director of a nonprofit agency to conduct an evaluation of a

<table>
<thead>
<tr>
<th>AEA Guiding Principles</th>
<th>CES Guidelines for Ethical Conduct</th>
<th>Ethical Principles for Evaluators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic inquiry</td>
<td>Evaluators should apply systematic methods of inquiry appropriate to the evaluation</td>
<td>Maximizing benefits Minimizing harms Balancing harms and benefits</td>
</tr>
<tr>
<td>Competence</td>
<td>Evaluators are to be competent in their provision of service</td>
<td>Minimizing harms</td>
</tr>
<tr>
<td>Integrity/honesty</td>
<td>Evaluators are to act with integrity in their relationships with all stakeholders</td>
<td>Being honest Keeping promises No conflicts of interest</td>
</tr>
<tr>
<td>Respect for people</td>
<td>Evaluators should confer with the client on contractual decisions, such as confidentiality, privacy, communication, and ownership of findings and reports</td>
<td>Free and informed consent Privacy and confidentiality Respect for vulnerable persons</td>
</tr>
<tr>
<td>Responsibilities for the general and public welfare</td>
<td></td>
<td>Procedural justice—ethical reviews of projects are fair, independent, and transparent. Distributive justice—persons are not discriminated against, and there is respect for vulnerable persons</td>
</tr>
</tbody>
</table>
major program that is delivered by the agency. The contract specifies that the evaluator will deliver three interim progress reports to the executive director, in addition to a final report. As the evaluator begins his or her work, he or she learns from several agency managers that the executive director has been redirecting money from the project budget for office furniture, equipment, and his or her own travel expenses—none of these being connected with the program that is being evaluated. In his or her first interim report, he or she brings these concerns to the attention of the executive director, who denies any wrongdoings, and makes it clear that the interim reports are not to be shared with anyone else. The evaluator discusses this situation with his or her colleagues in the firm in which he or she is employed and decides to inform the chair of the board of directors for the agency. He or she has broken his or her contract but has called on a broader principle that speaks to the honesty and integrity of the evaluation process.

In Appendix A, we have included a case that provides you with an opportunity to make a choice for an evaluator who works in a government department. The evaluator is in a difficult situation and has to decide what decision he or she should make, balancing ethical principles and his or her own well-being as the manager of an evaluation branch in that department. There is no right answer to this case. Instead, it gives you an opportunity to see how challenging ethical choice making can be, and it gives you an opportunity to make a choice and build a rationale for your choice. The case is a good example of what is involved in exercising deliberative judgment.

**Ethical Evaluation Practice**

Ethical behavior is not so much a matter of following principles as of balancing competing principles. (Stake & Mabry, 1998, p. 108)

Ethical practice in program evaluation is situation specific and can be challenging. The guidelines and principles discussed earlier are general. Sound ethical evaluation practice is circumstantial, much like sound professional judgment. Practice with ethical decision making is essential, dialogue being a key part of learning how ethics principles apply to practice and how principles feel subjectively.

How do we define ethical evaluation? Several definitions of sound ethical practice exist. Schwandt (2007, p. 401) refers to a “minimalist view,” in which evaluators develop sensitivity, empathy, and respect for others, and a “maximalist view,” which includes specific guidelines for ethical practice including “[recording] all changes made in the originally negotiated project plans, and the reasons why the changes were made” (AEA, 2004). Stake and Mabry (1998) define ethics as “the sum of human aspiration for honor in personal endeavor, respect in dealings with one another, and fairness in the collective treatment of others” (p. 99). Schweigert (2007) defines stated ethics in program evaluation as “limits or standards to prohibit intentional harms and name the minimum acceptable levels of performance” (p. 394). Ethical problems in evaluation are often indistinct, pervasive, and difficult to resolve with confidence (Stake & Mabry, 1998).

Although guidelines and professional standards can help guide the evaluator toward more ethical decisions, they have been criticized as lacking enforceability and
failing to anticipate the myriad situations inevitable in practice (Bamberger, Rugh, & Mabry, 2012)—hence the call for cultivating sound professional judgment (through reflective practice) in applying the principles and guidelines.

Like other professional judgment decisions, appropriate ethical practice occurs throughout the evaluation process. It usually falls to the evaluator to lead by example, ensuring that ethical principles are adhered to and are balanced with the goals of the stakeholders. Brandon, Smith, and Hwalek (2011), in discussing a successful private evaluation firm, describe the process this way:

Ethical matters are not easily or simply resolved but require working out viable solutions that balance professional independence with client service. These are not technical matters that can be handed over to well-trained staff or outside contractors, but require the constant, vigilant attention of seasoned evaluation leaders. (p. 306)

In contractual engagements, the evaluator has to make a decision to move forward with a contract or, as Smith (1998) describes it, to determine if an evaluation contract may be “bad for business” (p. 178). Smith goes on to recommend declining a contract if the desired work is not possible at an “acceptable level of quality” (Smith, 1998, p. 178). For internal evaluators, turning down an evaluation contract may have career implications. The case study at the end of this chapter explores this dilemma. Smith (1998) cites Mabry (1997) in describing the challenge of adhering to ethical principles for the evaluator:

Evaluation is the most ethically challenging of the approaches to research inquiry because it is the most likely to involve hidden agendas, vendettas, and serious professional and personal consequences to individuals. Because of this feature, evaluators need to exercise extraordinary circumspection before engaging in an evaluation study. (Mabry, 1997, p. 1, cited in Smith, 1998, p. 180)

Cultural Competence in Evaluation Practice

While issues of cultural sensitivity are addressed in Chapter 5, cultural sensitivity is as important for quantitative evaluation as it is for qualitative evaluation. We are including cultural competence in this section on ethics, as cultural awareness is an important feature of not only development evaluation where we explicitly work across cultures, but also virtually any evaluations conducted in our increasingly multicultural society. Evaluations in the health, education, or social sectors, for example, would commonly require that the evaluator have cultural awareness and sensitivity.

There is evidence of a growing sense of the importance and the relevance of acknowledging cultural awareness for evaluations. Schwandt (2007) notes that “the Guiding Principles (as well as most of the ethical guidelines of academic and professional associations in North America) have been developed largely against the foreground of a Western framework of moral understandings” (p. 400) and are often framed in terms of individual behaviors, largely ignoring the normative influences of social practices and institutions. The AEA
Guiding Principles for Evaluators include the following caveat to address the cross-cultural limitations of their principles:

These principles were developed in the context of Western cultures, particularly the United States, and so may reflect the experiences of that context. The relevance of these principles may vary across other cultures, and across subcultures within the United States. (AEA, 2004)

Schwandt (2007) notes that “in the Guiding Principles for evaluators, cultural competence is one dimension of a general principle (‘competence’) concerned with the idea of fitness or aptitude for the practice of evaluation” (p. 401); however, he challenges the adequacy of this dimension, asking “Can we reasonably argue for something like a cross cultural professional ethic for evaluators, and if so, what norms would it reflect?” (p. 401). Schwandt (2007) notes that the focus on cultural competence in evaluation has developed out of concern for “attending to the needs and interests of an increasingly diverse, multicultural society and the challenges of ensuring social equity in access to and quality of human service programs” (p. 401). In an imagined dialogue between two evaluators, Schwandt and Dahler-Larsen (2006) discuss resistance to evaluation and the practical implications for performing evaluation in communities. They conclude that “perhaps evaluators should listen more carefully and respond more prudently to voices in communities that are hesitant or skeptical about evaluation [. . . ] Evaluation is not only about goals and criteria, but about forms of life” (p. 504).

THE PROSPECTS FOR AN EVALUATION PROFESSION

In this chapter, we have emphasized the importance of acknowledging and cultivating sound professional judgment as part of what we believe is required to move evaluation in the direction of becoming a profession. In some professions, medicine being a good example, there is growing recognition that important parts of sound practice are tacitly learned, and that competent practitioners need to cultivate the capacity to reflect on their experience to develop an understanding of their own subjectivity and how their values, beliefs, expectations, and feelings affect the ways that they make decisions in their practice.

Some evaluation associations, the Canadian Evaluation Society (CES) being the most prominent example, have embarked on a professionalization path that has included identifying core competencies for evaluators and offering members the option of applying for a professional designation. Knowledge (formal education), experience, and professional reputation are all included in the assessment process conducted by an independent panel, and successful applicants receive a Credentialed Evaluator designation (CES, 2012b).

Other evaluation associations, with their emphasis on guidelines and standards for evaluation practice, are also embarking on a process that moves the field toward becoming more professional. Efforts are being made to identify core competencies (King et al., 2001), and discussions have outlined some of the core epistemological and
methodological issues that would need to be addressed if evaluation is to move forward as a profession (Bickman, 1997; Patton, 2008). The evaluation field continues to evolve as academic and practice-based contributors offer new ideas, critique each other’s ideas, and develop new approaches. Becoming more like a profession will mean balancing the norms of professional practice (core body of knowledge, ethical standards, and perhaps even entry to practice requirements) with the ferment that continues to drive the whole field and makes it both challenging and exciting.

Although many evaluators have made contributions that suggest we are moving toward making evaluation into a profession, we are not there yet. Picciotto (2011) concludes the following:

Evaluation is not a profession today but could be in the process of becoming one. Much remains to be done to trigger the latent energies of evaluators, promote their expertise, protect the integrity of their practice and forge effective alliances with well wishers in government, the private sector and the civil society. It will take strong and shrewd leadership within the evaluation associations to strike the right balance between autonomy and responsiveness, quality and inclusion, influence and accountability. (p. 179)

SUMMARY

Program evaluation is partly about learning methods and how to apply them. But, because most evaluation settings offer only roughly appropriate opportunities to apply tools that are often designed for social science research settings, it is essential that evaluators learn the craft of working with square pegs for round holes. Evaluators and managers have in common the fact that they are often trained in settings that idealize the applications of the tools that they learn. When they enter the world of practice, they must adapt what they have learned. What works is determined by the context and their experiences. Experience becomes the foundation not only of when and how to apply tools but, more important, the essential basis for interpreting the information that is gathered in a given situation.

Evaluators have the comparative luxury of time and resources to examine a program or policy that managers usually have to judge in situ, as it were. Even for evaluators, there are rarely sufficient resources to apply the tools that would yield the highest quality of data. That is a limitation that circumscribes what we do, but does not mean that we should stop asking whether and how programs work.

This chapter emphasizes the central role played by professional judgment in the practice of professions, including evaluation, and the importance of cultivating sound professional judgment. Michael Patton, through his alter ego Halcolm, puts it this way (Patton, 2008, p. 501):

Forget “judge not and ye shall not be judged.”

The evaluator’s mantra: Judge often and well so that you get better at it.

—Halcolm
It follows that professional programs, courses in universities, and textbooks should underscore for students the importance of developing and continuously improving their professional judgment skills, as opposed to focusing only on learning methods, facts, and exemplars. Practicing the craft of evaluation necessitates developing knowledge and skills that are tacit. These are learned through experience, refined through reflective practice, and applied along with the technical and rational knowledge that typically is conveyed in books and in classrooms. Practitioners in a profession begin to recognize that practice is much more messy than they were led to believe [in school], and worse, they see this as their own fault—they cannot have studied sufficiently well during their initial training. . . . This is not true. The fault, if there is one, lies in the lack of support they receive in understanding and coping with the inevitably messy world of practice. (Fish & Coles, 1998, p. 13)

Fish and Coles continue thus:

Learning to practice in a profession is an open capacity, cannot be mastered and goes on being refined forever. Arguably there is a major onus on those who teach courses of preparation for professional practice to demonstrate this and to reveal in their practice its implications. (p. 43)

The ubiquity of different kinds of judgment in evaluation practice suggests that as a nascent profession we need to do at least three things. First, we need to fully acknowledge the importance of professional judgment and the role it plays in the diverse ways we practice evaluation. Second, we need to understand how our professional judgments are made—the factors that condition our own judgments. Reflective practice is critical to reaping the potential from experience. Third, we need to work toward self-consciously improving the ways we incorporate, into the education and training of evaluators, opportunities for current and future practitioners to improve their professional judgments. Embracing professional judgment is an important step toward more mature and self-reflective evaluation practice.

Ethical evaluation practice is a part of cultivating sound judgment. Although national and international evaluation associations have developed principles and guidelines that include ethical practice, these guidelines are general and are not enforceable. Individual evaluators need to learn, through their reflective practice, how to navigate the ethical trade-offs in situations, understanding that appropriate ethical practice will weigh the risks and benefits for those involved.

**DISCUSSION QUESTIONS**

1. Take a position for or against the following proposition and develop a strong one-page argument that supports your position. This is the proposition: “Be it resolved that experiments, where program and control groups are randomly assigned, are the Gold Standard in evaluating the effectiveness of programs.”
2. What do evaluators and program managers have in common? What differences can you think of as well?

3. What is tacit knowledge? How does it differ from public knowledge?

4. In this chapter, we said that learning to ride a bicycle is partly tacit. For those who want to challenge this statement, try to describe learning how to ride a bicycle so that a person who has never before ridden a bicycle could get on one and ride it right away.

5. What is mindfulness, and how can it be used to develop sound professional judgment?

6. Why is teamwork an asset for persons who want to develop sound professional judgment?

7. What do you think would be required to make evaluation more professional, that is, have the characteristics of a profession?
Appendix A: Fiona’s Choice: An Ethical Dilemma for a Program Evaluator

Fiona Barnes did not feel well as the deputy commissioner’s office door closed behind her. She walked back to her office wondering why bad news seems to come on Friday afternoons. Sitting at her desk, she went over the events of the past several days and the decision that lay ahead of her. This was clearly the most difficult situation that she had encountered since her promotion to the position of director of evaluation in the Department of Human Services.

Fiona’s predicament had begun the day before, when the new commissioner, Fran Atkin, had called a meeting with Fiona and the deputy commissioner. The governor was in a difficult position: In his recent election campaign, he had made potentially conflicting campaign promises. He had promised to reduce taxes and had also promised to maintain existing health and social programs, while balancing the state budget.

The week before, a loud and lengthy meeting of the commissioners in the state government had resulted in a course of action intended to resolve the issue of conflicting election promises. Fran Atkin had been persuaded by the governor that she should meet with the senior staff in her department, and after the meeting, a major evaluation of the department’s programs would be announced. The evaluation would provide the governor with some post-election breathing space. But the evaluation results were predetermined—they would be used to justify program cuts. In sum, a “compassionate” but substantial reduction in the department’s social programs would be made to ensure the department’s contribution to a balanced budget.

As the new commissioner, Fran Atkin relied on her deputy commissioner, Elinor Ames. Elinor had been one of several deputies to continue on under the new administration and had been heavily committed to developing and implementing key programs in the department, under the previous administration. Her success in doing that had been a principal reason why she had been promoted to deputy commissioner.

On Wednesday, the day before the meeting with Fiona, Fran Atkin had met with Elinor Ames to explain the decision reached by the governor, downplaying the contentiousness of the discussion. Fran had acknowledged some discomfort with her position, but she believed her department now had a mandate. Proceeding with it was in the public’s interest.

Elinor was upset with the governor’s decision. She had fought hard over the years to build the programs in question. Now she was being told to dismantle her legacy—programs she believed in that made up a considerable part of her budget and person-year allocations.

In her meeting with Fiona on Friday afternoon, Elinor had filled Fiona in on the political rationale for the decision to cut human service programs. She also made clear what Fiona had suspected when they had met with the commissioner earlier that week—the outcomes of the evaluation were predetermined: They would show that key programs where substantial resources were tied up were not effective and would be used to justify cuts to the department’s programs.

Fiona was upset with the commissioner’s intended use of her branch. Elinor, watching Fiona’s reactions closely, had expressed some regret over the situation. After some
hesitation, she suggested that she and Fiona could work on the evaluation together, “to ensure that it meets our needs and is done according to our standards.” After pausing once more, Elinor added, “Of course, Fiona, if you do not feel that the branch has the capabilities needed to undertake this project, we can contract it out. I know some good people in this area.”

Fiona was shown to the door and asked to think about it over the weekend.

Fiona Barnes took pride in her growing reputation as a competent and serious director of a good evaluation shop. Her people did good work that was viewed as being honest, and they prided themselves on being able to handle any work that came their way. Elinor Ames had appointed Fiona to the job, and now this.

Your Task

Analyze this case and offer a resolution to Fiona’s dilemma. Should Fiona undertake the evaluation project? Should she agree to have the work contracted out? Why?

In responding to this case, consider the issues on two levels: (1) look at the issues taking into account Fiona’s personal situation and the “benefits and costs” of the options available to her and (2) look at the issues from an organizational standpoint, again weighing the “benefits and the costs.” Ultimately, you will have to decide how to weigh the benefits and costs from both Fiona’s and the department’s standpoints.

REFERENCES


Mabry, L. (1997). Ethical landmines in program evaluation. In R. E. Stakes (Chair), *Grounds for turning down a handsome evaluation contract.* Symposium conducted at the meeting of the AERA, Chicago, IL.


Chapter 12  The Nature and Practice of Professional Judgment