

To justify evaluative conclusions, evaluators follow both a general and a working logic.

Establishing Evaluative Conclusions: A Distinction Between General and Working Logic

Deborah M. Fournier

We say we can evaluate almost anything. In our attempt to do so, we commonly begin an evaluation with a question to answer or problem to solve: what are the effects of the student at-risk program statewide? Is centralized adult health care better than decentralized? In answering such questions, we as evaluators collect evidence and draw inferences in such a way as to build a strong, plausible argument that is meaningful to an often diverse audience (House, 1980, 1992). Ultimately evaluators reason their way to a concluding statement or series of evaluative conclusions as to the merit or worth of a program, product, or person.

In making evaluative conclusions about something, anyone can then ask for reasons supporting such claims. Our evidence and reasoning, as well as our conclusions, are subject to challenge and criticism. The decision whether to seriously believe in and act on the conclusions depends in part on the reasoning used to build the case.

As with all inquiry, a fundamental concern is how to go about building a strong, credible case: what is the reasoning process by which evaluative conclusions are established and supported? In what ways are evaluative conclusions justified?

The means to developing, strengthening, and clarifying reasoning that leads to legitimate evaluative conclusions is the crux of successful evaluation theory and practice. Evaluators are in the business of establishing the defensibility of empirical and normative claims about some phenomenon to clients

My thanks to Nick L. Smith, David R. Krathwohl, and Emily Roberston for their insightful comments on the development of the ideas presented in this chapter.

and other stakeholder groups. The kind of reasoning used to justify conclusions is of constant concern to evaluators “largely because the worry about the warrant for conclusions drawn from *any* inquiry will not wane” (Phillips, 1992, p. 119). Furthermore, “the reasoning process behind evaluative claims is vital because at any one time, the viewpoint that is the most objective is the one that currently is the most *warranted* or rational. . . . If we give up this distinction, if we hold that a biased or personally loaded viewpoint is as good as a viewpoint supported by carefully gathered evidence, we are undermining the very point of human inquiry. . . . If a shoddy inquiry is to be trusted as much as a careful one, then it is pointless to inquire carefully” (p. 68).

In the attempt to clarify reasoning used to establish evaluative claims, I propose the notions of general logic and working logic. This distinction is introduced as a way to assist evaluators in thinking about the reasoning process that they go through in arriving at evaluative conclusions, to bring out some of the important aspects of reasoning that may yet go unrecognized in day-to-day practice.

General Logic

There are different kinds of inquiry across practice areas, such as that which is found in law, medicine, and science. Common to each kind of inquiry is a general pattern of reasoning or basic logic that guides and informs the practice (Toulmin, 1964, 1984; McCarthy, 1973, 1979; Redding, 1989; Taylor, 1961). This basic logic provides practitioners with the rules for constructing and testing claims, and it specifies the basic conditions under which rationally motivated argumentation can take place. That is, it specifies to practitioners how someone would reason to justify his or her claims. In doing so, it also serves to distinguish one field of inquiry from other kinds of inquiry. Without a basic logic circumscribing the inquiry process, there is only a loose set of activities.

Evaluation is one kind of inquiry, and it, too, has a basic logic or general pattern of reasoning. This has already been put forth in the logic of product evaluation advanced by Michael Scriven (1980, 1981, 1990, 1993). This general logic of evaluation is as follows:

1. *Establishing criteria of merit.* On what dimensions must the evaluand do well?
2. *Constructing standards.* How well should the evaluand perform?
3. *Measuring performance and comparing with standards.* How well did the evaluand perform?
4. *Synthesizing and integrating data into a judgment of merit or worth.* What is the merit or worth of the evaluand?

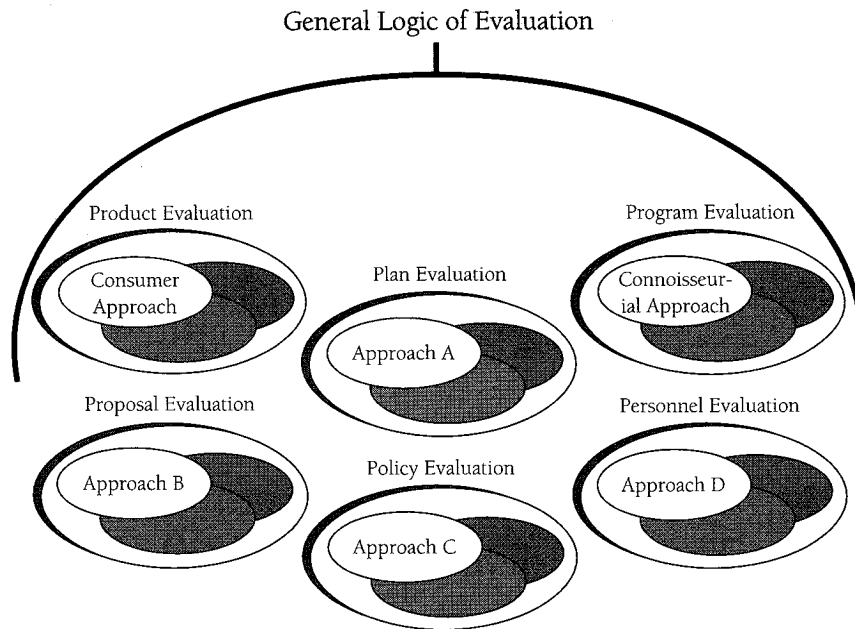
This general logic is also clearly evident in various philosophical discussions of value theory (see Hare, 1972; Rescher, 1969; Taylor, 1961). To evaluate anything means to assess the merit or worth of something against criteria and standards. The basic logic explicated by Scriven reflects what it means when we use the term *to evaluate*.

The general logic can be found across various instances of the evaluation inquiry process. For example, the numerous evaluation approaches developed

by theorists vary from one another in many details, yet I find that they do share this common logic (Fournier, 1993). Connoisseurial, judicial, pluralistic, and goal-free approaches are all instances of evaluation practice (see Alkin and Ellett, 1990; Madaus, 1989; McLaughlin and Phillips, 1991). What counts as criteria or evidence and how evidence is weighed varies from one approach to another, yet all follow the pattern of evaluative reasoning noted in the four steps.

In turn, each evaluation approach belongs (or can belong) to a field of professionalized evaluation, namely, product evaluation, program evaluation, personnel evaluation, policy evaluation, proposal evaluation, and plan evaluation, to name a few (Scriven, 1991). Thus this general logic of evaluation is one that also overarches all fields within evaluation. It is the basic reasoning that specifies what it means to evaluate something, an activity logically distinct from, say, biomedical research. In other words, it specifies the game and the rules of the game that one is playing when conducting an evaluation in any field. If someone says that he or she is doing an evaluation, then he or she must be setting criteria and standards, measuring the evaluand along these lines, and synthesizing the information into a final judgment about the merit or worth of the evaluand. In turn, these four aspects of general logic are used to critique practice. For example, an evaluation would be objectionable if it determined that something had merit but could not provide criteria when required to do so. So the general logic helps distinguish evaluation from nonevaluation types of inquiry; it plays a critical role in defining and establishing professional identity and subsequent developments. (See Figure 2.1.)

Figure 2.1. Illustration of How General Logic Overarches All Fields and Approaches Within Evaluation



The evaluation inquiry game, to use Smith's term (see Chapter One), comes from an evaluative point of view in making evaluative types of claims by following the rules of this general logic in answering the following question: what is it to rationally assert the merit or worth of X? Although the various approaches found within the different fields follow this logic, they answer this question differently.

Working Logic

Subsumed under the general logic are many individual working logics. Working logic is the variation in detail in which the general logic is followed when conducting an evaluation. In other words, what varies across approaches is what or how criteria are identified, what or how standards are constructed, how performance is measured, and how data are synthesized. For example, in a consumer approach to product evaluation a source of criteria is the properties inherent in the product (Scriven, 1980, 1981). In contrast, others have suggested that a source of criteria for evaluating social programs is what stakeholders value in a particular program (Shadish, Cook, and Leviton, 1991) or values personally held by an expert (Eisner, 1989, 1991).

Working logic is a term taken from the work of Stephen Toulmin (1964, 1972) and is tied to his characterization of disciplines as rational enterprises. Using Abraham Kaplan's (1964) distinction between reconstructed logic and logic-in-use, working logic is the logic-in-use found in everyday practice to establish and justify evaluative claims.

Each evaluation approach has its own working logic. Working logic guides and informs evaluators about how to reason to justify conclusions using a specific evaluation approach. By analogy, general logic is the strategy and working logic is the specific tactic. To justify conclusions made in evaluation, evaluators follow both forms of logic. Evaluators all follow the same general logic and employ a particular instance of it, a working logic. And theorists advance both kinds of logic in the approaches to practice they advocate.

There are two different ways of thinking about working logic in evaluation. First, working logic can be conceptualized as a set of four parameters that circumscribe the boundaries in which the reasoning process takes place: problem, phenomenon, question, and claim. Second, working logic can be conceptualized as an argument structure that reveals the reasoning pattern used to support conclusions.

Working Logic as Four Parameters: Problem, Phenomenon, Question, and Claim

Across evaluations there are differences in how the four steps of the general logic are followed. In step 1, the set of criteria identified can vary from evaluation to evaluation depending on the kind of evaluand under study. To identify criteria against which the evaluand will be measured in a particular study,

a clear understanding of the phenomenon is first needed (Scriven, 1980, 1981). This does not simply mean that an evaluator must know what he or she is evaluating but rather that he or she have a thorough understanding of its parts, organization, or structure; how it works; and how it relates to the larger context. Without this deep-seated feel for the phenomenon, the evaluator may select inappropriate criteria and omit critical ones. This is a serious problem; criteria can make or break an evaluation because they establish the basis for evaluative conclusions and thus directly affect the validity of claims.

Criteria selection is also influenced by the kind of questions being asked in the inquiry. A clear formulation of the questions is needed to identify meaningful criteria. Understanding what kind of phenomenon is involved in a certain evaluation and what kind of question is being asked influences step 1 of the general logic. This, of course, influences the standards that are constructed in step 2. In turn, the kind of questions posed will obviously influence the kind of conclusions or claims drawn in step 4.

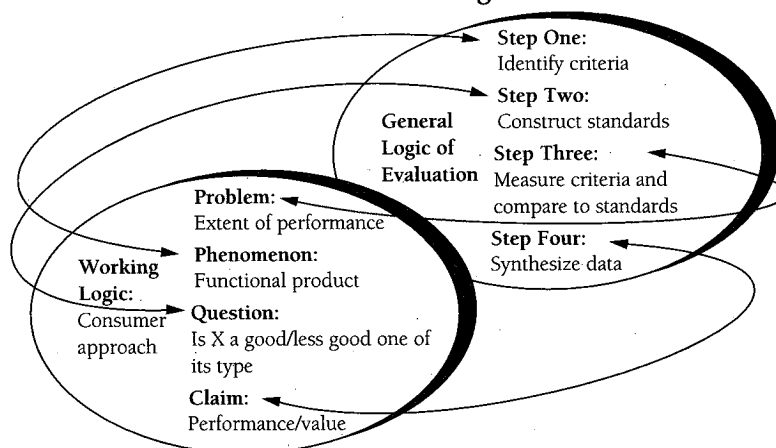
The kind of phenomena, questions, and claims are shaped by the kind of problem addressed in a particular evaluation. Also, in measuring the evaluand, the kind of problem being investigated influences how measures are to be taken (step 3). For instance, in a connoisseurial approach the problem is to describe the unique qualities of a program. Expert observation of participants is an appropriate means for this. However, such direct observation may be inappropriate for getting at causal relationships when dealing with problems of intervention effectiveness.

Thus differences in how the general logic is followed are due to the particular type of problems addressed by an evaluation, which in turn influences how the phenomenon under study is (or can be) defined, the kinds of questions raised about the phenomenon, and the kinds of claims that are ultimately made in the inquiry. These are the four parameters that circumscribe the boundaries of a given working logic: problem, phenomenon, question, and claim. (See Figure 2.2. Note that the parameters and steps are obviously interwoven and thus do not fit neatly in any one place. For the sake of categorizing the variation within general logic, the figure loses the sense of interaction among parameters.)

These parameters of working logic set the foundation for building an argument that works to establish and support conclusions (that is, reasoning is directed toward this particular set of parameters). Each evaluation approach has its own set of values for these parameters. Every theorist advances not only a particular approach but also a particular set of parameters—a certain way to establish evaluative conclusions. And because evaluation practice varies widely along these four parameters, evaluation practice can be viewed as consisting of a profusion of individual working logics.

For example, the working logic of a consumer approach to product evaluation (Scriven, 1981, 1990, 1993) focuses on determining the extent of performance (the problem) of functional products such as computers or cars (the phenomenon). The goal is to establish a performance/value conclusion (the

Figure 2.2. The Relationship Between Working Logic and General Logic



claim) that answers questions such as Is *X* a good one of its type? Is this *X* good or less good when compared with other *X*'s? (the question).

In contrast, a connoisseurial approach to program evaluation (Eisner, 1989, 1991) focuses on describing the unique qualities of a program (the problem) that is defined as a collection of qualities (the phenomenon). Its goal is to establish a descriptive or value conclusion (the claim) that answers questions such as What does it feel like to be in this program (as perceived by this expert)? What are the qualities that make this program good or less than good?

Still, a causal approach to program evaluation focuses on determining the intervention effectiveness of a program (the problem) that is defined as a set of treatment-outcome relationships (the phenomenon). The relationship has been structured in such a way as to ameliorate a particular educational or social problem. The goal is to establish a causal or value conclusion (the claim) that answers questions such as Is *A* more effective than *B* in producing *X*? Does program *A* cause more of *X* than program *B*? (the question) (Cook, 1991; Cook and Campbell, 1979). This approach obviously differs along all four parameters of working logic when compared with a consumer approach to product evaluation or a connoisseurial approach to program evaluation. (See Figures 2.3 and 2.4.)

Notice in these examples that the connoisseurial and causal approaches both deal with programs. Although on one level both approaches deal with the same type of phenomenon, namely programs, the phenomenon is different in that the approaches differ in the way programs are defined or conceptualized; that is, each has a different social construction. A causal approach defines a program as a set of treatment-outcome relationships. A connoisseurial approach defines a program as a collection of unique qualities. The two approaches therefore deal with different phenomena (and ask different questions of them). These approaches use two different kinds of working logic within the field of program evaluation.

Figure 2.3. Parameters of Different Working Logics

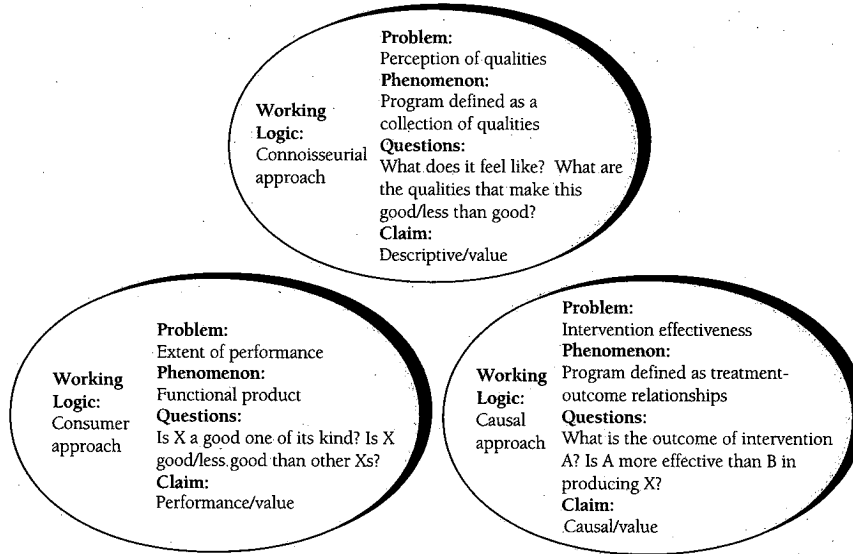
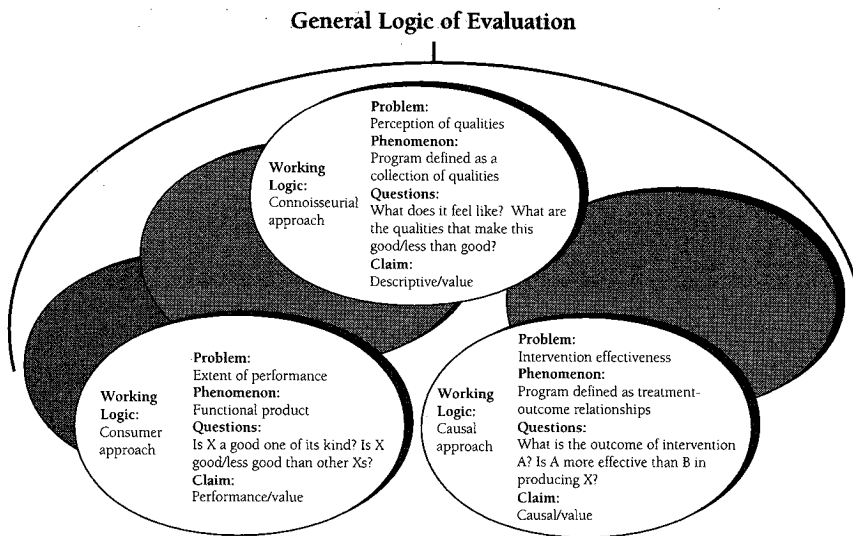


Figure 2.4. Individual Working Logics Subsumed Under the General Logic



Typically, evaluation approaches (that is, working logics) are developed in one field of evaluation. But it may be fruitful to employ an approach in other fields. For example, a causal approach to program evaluation might be used in product evaluation. A product could be defined as a set of treatment-outcome relationships, questions concerning cause and effect could be asked about products (such as prescription and nonprescription drugs or durable medical equipment such as wheelchairs, walkers, and the like). Yet it might not make sense to use the same approach in, say, policy evaluation. Or a consumer approach to product evaluation might be used in program evaluation. A program could be defined as a functional product; that is, program goals could be viewed as the functional properties of a program. In fact, we are all familiar with the long-standing objectives-based or goal-based approaches to evaluation that focus solely on evaluating the extent to which program objectives or goals have been attained. So parameters can be examined for their applicability in various fields. They are the ways in which we would expect approaches to differ.

One last point is that how the phenomenon is defined (that is, socially constructed) is important because it influences the source or locus of values from which criteria are selected (step 1 of the general logic). In turn, criteria selection affects the validity of conclusions because it influences the reasoning used in establishing them. The reasoning is affected because the source of the criteria commits us to look for certain kinds of evidence and to appeal to certain kinds of warrants in order to justify resulting claims. In other words, how evaluators reason toward evaluative judgments depends on how value (criteria) is defined. Thus, the burden of justification rests on the criteria.¹

For example, in a connoisseurial approach to program evaluation the goal is to reveal the qualities of the program as perceived by an expert so that others can see and feel what it is like to be in that program, the program being defined here as a set of qualities (Eisner 1989, 1991). The approach reasons toward a judgment of merit or worth concerning some of the qualities of a program. Given that the source of criteria is personally held values of an expert, then the kind of evidence an evaluator would be expected to look for would be a statement of these values. To infer that a program is good according to this expert is to then expect an evaluator to appeal to warrants that assert that this expert's judgment is reliable. To warrant claims in this approach is to affirm the credibility and reliability of the expert.

In contrast, in a pluralistic approach to program evaluation, the phenomenon of interest is, again, a program, but it is defined as a set of values and expectations that are held for a program by its stakeholders (Shadish, Cook, and Leviton, 1991; Stake, 1989, 1991). The approach reasons toward a description of a program so that decision makers and other stakeholders can vicariously experience what it is like to be involved in that program. In this approach it is primarily stakeholder values that are required to evaluate the merits of a program. Given that the source of criteria is stakeholder values, then the kind of evidence an evaluator would be expected to look for would

be a statement of these values. To infer program impact based on the evidence of values is to expect an evaluator to appeal to warrants that assert that stakeholder values is motive enough for accepting the conclusion because the program aims to serve stakeholders, and they are in the best position to know what that service should be like. (See Table 2.1 for additional examples.)

Therefore the first step of the general logic and the parameter of phenomenon is the most crucial concern for evaluators. Seeing theorists as developing alternative working logics for evaluation practice, thus committing them to take somewhat different stances toward the making and justifying of evaluative claims, suggests that perhaps the various debates about what constitutes good evaluations are often really hidden differences about how the phenomenon is viewed.

Working Logic as a Reasoning Pattern

A second way to conceptualize working logic is as an argument structure that makes concrete the reasoning pattern by which conclusions are justified. Whereas the parameters define the scope of a particular working logic, this

Table 2.1. Comparison of Different Definitions for the Same Phenomenon

<i>Evaluation Approach</i>	<i>Phenomenon of Interest</i>	<i>Source of Criteria (Locus of Values)</i>	<i>Evidence (Foundation for a Claim)</i>	<i>Warrant (authorizes Inference)</i>
Connoisseurial/critic approach to program evaluation	Program defined as set of qualities identifiable by an expert	Personally held values of an expert	Expert values	Expert is reliable and credible
Pluralistic approach to program evaluation	Program defined as set of values held by stakeholders	Stakeholder values	Stakeholder values and their connection to impact	Stakeholder values reflect what is desirable and important
Consumer approach to product evaluation	Functional product	Properties inherent in the product and consumer use	Properties and their connection to extent of performance	Accepted meaning of the word (such as car or watch)
Goal-free approach to program evaluation	Program defined as a means of meeting needs	Consumer needs	Needs and their connection to program effects	Needs accepted as necessary requirements for existence
Causal approach to program evaluation	Program defined as set of treatment-outcome relationships	Dependent variables in goals or research literature	Relationships among variables	Relationships were identified under reliable methods

second conceptualization lays out the reasoning pattern used to justify conclusions within given parameters.

Stephen Toulmin (1964) examined different types of inquiry and found that inquiry is best characterized as the building of a defensible argument. He identified six logical features that are common to all kinds of inquiry. All these features work together to support and justify conclusions resulting from a process of inquiry:

Claims that conclude what is to be taken as acceptable and legitimate

Evidence, that is, the facts forming the basis for the claim

Warrants that legitimate the inferences drawn from evidence by appeal to some authority

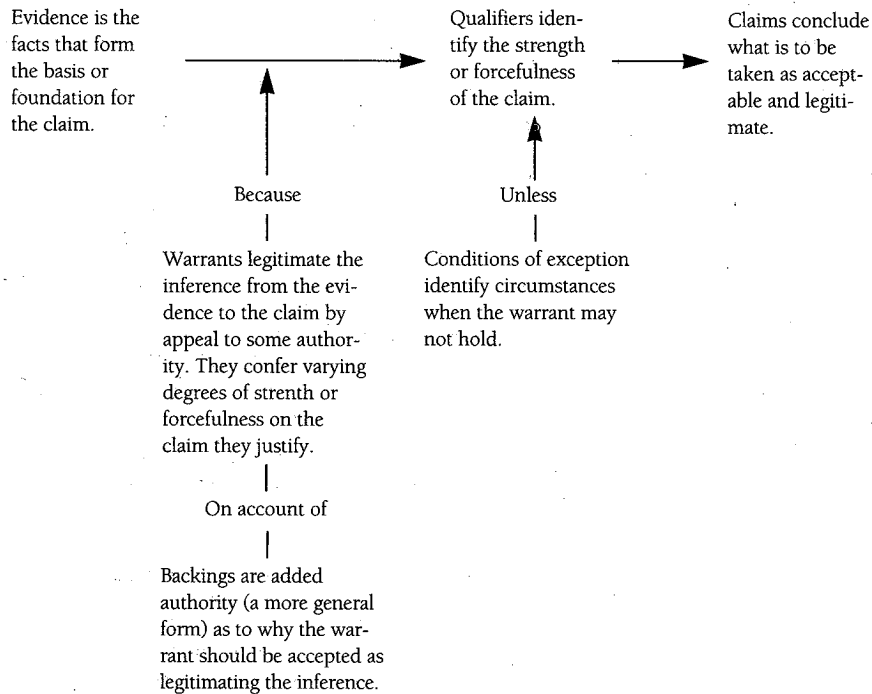
Backings that support the warrant by appeal to some more general authority

Conditions of exception that point out circumstances when the warrant may not hold

Qualifiers that identify the strength of the claim (see Figure 2.5).

This argument structure can be a useful tool in mapping out the reasoning process in any evaluation. In earlier work, I have explained each feature in detail and applied it to some of the evaluation approaches in order to clarify

Figure 2.5. Six Main Logical Features Common to All Inquiry



the merit of each approach. (See Fournier 1992; Fournier and Smith, 1993.) In this work I applied the features and used them to explicate and summarize the overall reasoning of a particular evaluation approach. But typically an overall argument contains a collection of subarguments linked together in various configurations. The structure is useful in mapping underlying subarguments because more than single pieces of evidence, warrants, backings, conditions of exception, and qualifiers are usually offered to support and legitimate claims made in evaluations. The structure is a useful concept to examine reasoning at multiple levels.

In the previous section I stated that parameters of working logic can vary across evaluation approaches. Although one parameter, the type of claim, may vary across different working logics, all claims are similar in that they depend on five logical features of an argument: evidence, warrant, backing, conditions of exception, and qualifier. Given space constraints, I will discuss only one of these features in this chapter, the warrant, to suggest one way in which this tool can be useful.

Warrant-Establishing Versus Warrant-Using

The warrant is what legitimates the inferences drawn from evidence by appeal to some kind of authority. The warrants on which claims rely vary from one area of practice to another. Certain kinds of warrant are deemed acceptable by members within an area of practice. For instance, in science claims are made about causation by appeal to the randomization of samples. Appeals to another authority such as "because I saw it with my own eyes" or "because God revealed it to me" would not be acceptable warrants for causal claims, although such appeals might be most appropriate in other areas (for example, law or religion).

To continue with the science example, in an experimental study where randomization is possible, the warrantability of causal claims is grounded in sampling theory. Sampling theory tells us that when experiments are randomized, influences on the dependent variable are dispersed equally over treatment groups, resulting in a no-difference pretest comparison (that is, randomization achieves comparative equivalence). Comparison groups are employed and pre-post measures are collected. Randomization, comparison groups, and pre-post measures serve as warrants to establish and legitimate the making of causal inferences. Sampling theory goes further and even indicates the strength or force of such warrants by relating sample sizes to variances. Samples should be of a certain size relative to variance to strengthen the claim.

In studies in which randomization is not possible, the warrantability of claims changes. Because of the lack of randomization, and subsequently the associated warrantability that sampling theory provides, the key to strong, defensible claims in quasi-experimentation is the use of pattern matching of causal predictions to the obtained data (verification) and the ruling out of alternative explanations (falsification) (Campbell and Stanley, 1963; Cook,

1993; Cook and Campbell, 1979). As in experimental studies, pre-post measures and comparison groups are used in facilitating causal inferences. However, warrantability is grounded in verification and falsification, more specifically, point specificity (for example, pattern matching in time series and regression discontinuity designs) and in the complexity of causal explanations (for example, multivariate). Point specificity means that if the specific point where an anticipated effect will be observed is clearly identified ahead of time and there is change after this treatment point, then it can be inferred that the treatment had impact. The value of point specificity is that it strengthens the validity of claims. As regards complex causal explanations, the more complex the causal implications, the less likely it is that other explanations exist. The value of complex causal patterns is that they reduce alternative explanations, thereby strengthening claims. As with sampling theory, warrant in quasi-experimental studies also indicates to some extent the strength or force of its warrant. The more alternative explanations that can be ruled out, the stronger the claim.

The reasoning in experimentation is strong because it is based on the long-standing and highly developed area of sampling theory. Such warrant is, for the most part, uncontested and generally accepted as a given. Reasoning that uses this type of warrant is a warrant-using argument (Toulmin, 1964). However, many evaluation approaches are not of this kind. Rather, evaluation approaches tend to be warrant-establishing arguments (Toulmin, 1964). This term refers to warrants that are not uncontested and must be constructed and tried out by application in a number of cases. Here the yet unestablished warrant, as well as the claim, must be legitimated. It is the difference between the "taking of a journey along a railway already built and the building of a fresh railway" (Toulmin, 1964, p. 120). In evaluation, the various approaches seek to legitimate both evaluative claims and warrants.

The development of a quasi-experimental approach is a familiar example of such a warrant-establishing argument in evaluation. "The last thirty years of systematic program evaluation have witnessed two sets of changes . . . in the theory of quasi-experimentation. . . . They concern changes in the intellectual warrant justifying the molar descriptive causal conclusions" (Cook, 1991, p. 116).

In developing quasi-methods for causal inquiry, Campbell began by identifying alternative explanations or causes that random assignment rules out (such as history and maturation). The warrant for causal inferences could be facilitated through falsification. Shortcomings in this warrant led Campbell and Cook to make modifications. One obvious weakness is that the list of alternative causes presumes that it is a comprehensive one. Recognizing this as a weakness, Campbell and Cook later added verification to the original falsification warrant. "Matching specific or complex causal predictions with obtained data became the order of the day" (Cook, 1991, p. 120).

Notice that the warrant was tried out by applying it in a number of cases. Campbell generated his first list of alternative explanations or causes based on

his own experiences. Later he identified other new threats through the “reflection of practicing social scientists who had independently identified new threats while on the job” (Cook, 1991, p. 118).

The point is that Campbell and Cook have made outstanding progress in developing the warrantability of causal claims in quasi-experimentation over the past two decades (see Campbell and Stanley, 1963; Cook, 1991, 1993; Cook and Campbell, 1979; Overman, 1988). What contributed to their success, in part, is the fact that they concentrated their efforts at a detailed level. They focused intently on a certain kind of claim (causal effectiveness) and one kind of phenomenon (social intervention programs) conceptualized one way (typically defined as treatment–outcome relationships). Can we get to such an understanding with other approaches used in evaluation practice? Can we better understand and strengthen the different kinds of warrant-establishing arguments in evaluation practice? I believe that the distinction between general and working logic is one means to that end, for it illuminates and characterizes what Campbell and Cook did implicitly so that it can be replicated.

Implications for Practice and Evaluation Theory

New Way to View the Various Evaluation Approaches. The notions of general and working logic are useful to evaluators because they provide a new way to view the various approaches to evaluation. Traditionally, we have viewed evaluation approaches as outlining the purpose of the evaluation (such as formative, summative, or critical), the role of the evaluation (for example, to service management or policy), the role of the evaluator (for example, as facilitator or change agent), the role of the stakeholder (for example, as participatory or representative), the methodological orientation (quantitative or qualitative), the ideological orientation (such as managerial or consumerist), and so forth. This view has resulted in debates such as whether to allow stakeholders to participate in the evaluation, whether summative evaluation is better than formative, whether to include stakeholder views in an evaluation so as to improve the utilization of results, whether to describe or prescribe values, and so on.

These have been informative debates, but they do little to help us with the specifics by which evaluators can legitimate conclusions to stakeholders. Given this understanding of general and working logic, we can now look at these approaches as advancing reasoning on how to construct a particular argument and add other questions to debate. In addition to asking, “What is the purpose of an evaluation?” we can ask, “What question or problem is the evaluation probing?” In addition to asking, “What is the role of the evaluator?” we can ask, “What kinds of claims is the evaluator trying to make?” and “What kinds of warrants will support or weaken such claims?” The general/working logic distinction can help evaluators generate a long list of practical questions about the reasoning used to establish evaluative conclusions in any approach to practice—questions to examine existing approaches advanced by theorists and to raise against approaches put together in everyday practice. The distinction

would also be beneficial in expanding how students typically study evaluation approaches in graduate programs around the country.

Illuminates How Evaluative Reasoning Is Influenced. General/working logic is valuable in helping evaluators strengthen the reasoning process they will go through in drawing evaluative conclusions in day-to-day practice because it illuminates the ways in which reasoning is influenced (for example, phenomenon or claim). It makes clear the different features that need to be considered carefully when drawing conclusions in evaluation (such as evidence and warrant). It also shows several ways in which conclusions are open to challenge and criticism. In doing so, it provides evaluators with a series of critical questions by which to examine their practice: “what kind of claim(s) am I trying to make in this evaluation? What authority source does the warrant need to appeal to? How strong is the authority source in supporting inferences drawn from evidence? What conditions would undermine the warrant? How forceful is (or will be) the resulting conclusion?”

In addition, the reasoning pattern presented earlier in Figure 2.5 interconnects such questions in pointing to broader issues. For example, the evidence–warrant–backing relationship that is located on the left side of the argument structure in Figure 2.5 can be viewed as a means for “ruling in” support, and the conditions of exception–qualifier relationship represented on the right side can be viewed as a means for “ruling out” alternative explanations. Using this reasoning pattern, evaluators can ask questions about how an evaluation approach rules in support that strengthens the conclusion and rules out circumstances that weaken the conclusion.

Provides a Practical Concept for Studying Practice and Theory. General/working logic is a practical concept that is useful for studying and comparing what is implemented in practice and proposed in theory because it schematizes evaluative reasoning. It provides a standard means, a unifying construct, for making comparisons across approaches in various fields and settings that have not been easily made in the past. This is important because the “separation of evaluation into distinct subareas, such as program evaluation, product evaluation, educational evaluation, medical evaluation, and others has impeded the development of evaluation theory because similarities across areas are not apparent” (House, 1993, p. 89).

The concept also makes it easier to analyze and critique what is practiced in the field and what is proposed in theory. For instance, the warrant appeals to some authority that legitimizes the inferences drawn from evidence. Examining the kinds of warrants used in practice reveals the nature of the rationality of evaluation as experienced by practitioners. This provides insight into the nature of inferential thinking used by evaluators across diverse settings. Through analysis, progress on the strengthening of the warrantability of evaluative conclusions can be made. Still, the parameters of working logic are ways in which we would expect approaches to vary; thus we can make comparisons between approaches along these four aspects. For example, we might ask, “In what ways is a product the same as a program or a policy? In what ways do

phenomena influence the way evaluators justify claims?" The development of the features that describe general/working logic enables us to raise new questions about logical practice. In doing so, we may become clearer about what it means to evaluate something or someone, and what that might mean in various contexts.

Further, having a standard means for comparison could move us away from unproductive comparisons and toward the realization that each approach has its own merit. Rather than saying things such as, "A causal approach has a strong working logic and a connoisseurial approach has a weak one," we can instead ask ourselves questions such as, "If one is making a claim about how students perceive and make meaning, then what kind of warrant would best support inferences drawn from the evidence? What has this approach not considered? What are the actual merits of the approach in light of the kind of claim it seeks to make? What is the best that can be expected of such an approach?" And because warrants confer various degrees of force on the conclusion(s), we are able to examine and categorize types of conclusions and their associated degrees of forcefulness. What kinds of warrants are commonly employed in, for example, a connoisseurial approach to evaluation or in a pluralistic approach? What are the merits of each evaluation approach? What is achievable in one and not in another?

Accounts for Differences Between Diverse Positions. Working logic is valuable in that it is able to account for differences among the diverse positions advanced by theorists and practitioners. For example, it can account for the theoretical incompatibility, noted by Shadish, Cook, and Leviton (1991), in applying the logic of product evaluation, as advanced by Scriven, to program evaluation.

Shadish and others (1991) conducted a comprehensive review and critique of Scriven's scholarship. They contend that the logic of evaluation developed through the study of product evaluation does not fit well when applied to program evaluation (see pp. 83–101). To determine whether the logic fits, I closely examined Scriven's work in product evaluation (Fournier, 1993).

The difficulties in using the logic of product evaluation in program evaluation are a result of blurring the distinction between general and working logic. The objections of Shadish and others are really problems with the variations of detail in which the general logic is followed. In other words, their difficulties are in applying one working logic to another working logic, not with applying the general logic of evaluation to another field, nor with applying a working logic developed in one field to another field. Their difficulties are in using the working logic of a consumer approach to product evaluation as the working logic of a particular pluralistic approach to program evaluation.

Given that an individual working logic is specific to a particular approach to evaluation (that is, particular parameters and reasoning processes), working logic would be expected to transfer to another field if one is willing to accept the parameters of a given working logic. For instance, the working logic of a consumer approach to product evaluation proposed by Scriven can transfer to

program evaluation if one is willing to accept the way in which the phenomenon is defined, the kind of questions posed, and so on. A goal-based approach to program evaluation uses this working logic without conflict because it accepts such parameters—it subscribes to conceptualizing a program as having functional properties such as products (program goals). Shadish and others have difficulty with applying this logic because they subscribe to different parameters; they define the phenomenon differently, ask different questions, and so on. But they do agree that the general logic fits across various fields of evaluation. Shadish and others are correct in stating that evaluation logic is not the same for everything, but it is at the level of working logic that such differences arise. The seemingly logical misfit between the working logic of a consumer approach to product evaluation and program evaluation can be accounted for through the conceptualization of two kinds of logic.

Clarifies Professional Identity. When a particular field of inquiry does not have a clear, collective understanding of the nature of the enterprise, it develops loosely and diffusely (Toulmin, 1972). When an area of inquiry lacks an agreed-on understanding about the nature of its enterprise, “the task of arriving at ‘rational’ judgments is liable—for understandable reasons—to be subtle and debatable” (p. 396). The general logic of evaluation clarifies and defines the nature of the enterprise so that everyone involved is playing the same game.

General logic serves to distinguish evaluation from nonevaluation. In doing so, it clarifies what it means to evaluate something. On this basis it is possible to establish professional identity for the field and a self-concept for evaluators and to shape future developments. This is true because it circumscribes a substantive area of inquiry and orients it from a certain point of view in terms of the perceived nature of its particular enterprise, the nature of the phenomenon undertaken in study, and the nature of its concepts in terms of how members interpret that experience. Professional activities (intellectual and practical) are directed by its goals (well-defined or less specific) and guided by its logic. The discrepancy between what a field aims to achieve concerning some particular phenomenon (its goals) and what it actually can achieve (existing procedures and methods) determines the kinds of problems it seeks to address and the kind of methodological and theoretical developments that need to take place.

Note

1. The warrant is what legitimates the inference one makes when moving from the evidence to the claim. It is the way in which the evidence is connected to the claim. For instance, I state that the new musical playing at the theater is good (claim). To support my claim, I add that I read that it received a good review in last night’s newspaper (evidence on which the claim is based). My friend questions me and asks how can I say that. I am not being asked to produce more evidence, but rather to legitimate the inferential leap from the evidence to the claim. I am being asked to legitimate the inference that the musical is good because the newspaper reported it as such. I respond by saying that the reviewer is well-

respected by the performing arts community and that her judgment is highly valued and reliable (warrant). In this example, note that the warrant appeals to expertise. In establishing claims, warrants appeal to different types of authority (for example, because it is the law, because the physician said so, or because A is known to cause B).

References

- Alkin, M. C., and Ellett, F. S. "Development of Evaluation Models." In H. J. Walberg and G. D. Haertel (eds.), *The International Encyclopedia of Educational Evaluation*. Elmsford, N.Y.: Pergamon Press, 1990.
- Campbell, D. T., and Stanley, J. C. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin, 1963.
- Cook, T. D. "Clarifying the Warrant for Generalized Causal Inferences in Quasi-Experimentation." In M. C. McLaughlin and D. C. Phillips (eds.), *Evaluation and Education: At Quarter Century*. Chicago: University of Chicago, 1991.
- Cook, T. D. "A Quasi-Sampling Theory of the Generalization of Causal Relationships." In L. B. Sechrest and A. G. Scott (eds.), *Understanding Causes and Generalizing About Them*. New Directions for Program Evaluation, no. 57. San Francisco: Jossey-Bass, 1993.
- Cook, T. D., and Campbell, D. T. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin, 1979.
- Eisner, E. W. "Educational Connoisseurship and Criticism." In G. F. Madaus, M. Scriven, and D. L. Stufflebeam (eds.), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston: Kluwer-Nijhoff, 1989.
- Eisner, E. W. "Taking a Second Look: Educational Connoisseurship Revisited." In M. C. McLaughlin and D. C. Phillips (eds.), *Evaluation and Education: At Quarter Century*. Chicago: University of Chicago, 1991.
- Fournier, D. M. "Clarifying the Merits of Argument." In J. Zhao and M. A. Stahl (eds.), *Proceedings of the 1992 Edward F. Kelly Conference*. Albany, N.Y.: Evaluation Consortium, School of Education, SUNY at Albany, 1992.
- Fournier, D. M. "Reasoning in Evaluation: A Distinction Between General and Working Logic." Unpublished doctoral dissertation, Syracuse University, 1993.
- Fournier, D. M., and Smith, N. L. "Clarifying the Merits of Argument in Evaluation Practice." *Evaluation and Program Planning*, 1993, 17 (1), 315-323.
- Hare, R. M. "What is a Value Judgment?" In P. W. Taylor (ed.), *Problems of Moral Philosophy: Introduction to Ethics*. Encino, Calif.: Dickenson, 1972.
- House, E. R. *Evaluating with Validity*. Newbury Park, Calif.: Sage, 1980.
- House, E. R. "Integrating Quantitative and Qualitative." Paper presented at the American Evaluation Association Conference, Seattle, Nov. 1992.
- Kaplan, A. *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandler, 1964.
- McCarthy, T. A. "A Theory of Communicative Competence." *Philosophy of the Social Sciences*, 1973, 3, 135-156.
- McCarthy, T. A. *Communication and the Evolution of Society: Jürgen Habermas*. Toronto: Beacon, 1979.
- McLaughlin, M. C., and Phillips, D. C. *Evaluation and Education: At Quarter Century*. Chicago: University of Chicago, 1991.
- Madaus, G. F., Scriven, M., and Stufflebeam, D. L. *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston: Kluwer-Nijhoff, 1989.
- Overman, E. S. (ed.). *Methodology and Epistemology for Social Science: Selected Papers, Donald T. Campbell*. Chicago: University of Chicago, 1988.
- Phillips, D. C. *Social Scientist's Bestiary: A Guide to Fabled Threats to, and Defenses of, Naturalistic Social Science*. Elmsford, N.Y.: Pergamon Press, 1992.
- Redding, P. "Habermas' Theory of Argumentation." *Journal of Value Inquiry*, 1989, 23, 15-32.

- Rescher, N. *Introduction to Value Theory*. Englewood Cliffs, N.J.: Prentice Hall, 1969.
- Scriven, M. *The Logic of Evaluation*. Inverness, Calif.: Edgepress, 1980.
- Scriven, M. "Product Evaluation." In N. L. Smith (ed.), *New Techniques for Evaluation: New Perspectives in Evaluation*. Newbury Park, Calif.: Sage, 1981.
- Scriven, M. "The Evaluation of Hardware and Software." *Studies in Educational Evaluation*, 1990, 16, 3–40.
- Scriven, M. *Evaluation Thesaurus*. (4th ed.) Newbury Park, Calif.: Sage, 1991.
- Scriven, M. "Evaluation and Critical Reasoning: Logic's Last Frontier?" In R. Talaska (ed.), *Critical Reasoning in Contemporary Culture*. Albany: State University of New York, 1993.
- Shadish, W. R., Cook, T. D., and Leviton, L. C. *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, Calif.: Sage, 1991.
- Stake, R. E. "Program Evaluation, Particularly Responsive Evaluation." In G. F. Madaus, M. Scriven, and D. L. Stufflebeam (eds.), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston: Kluwer-Nijhoff, 1989.
- Stake, R. E. "Retrospective on 'The Countenance of Educational Evaluation.'" In M. C. McLaughlin and D. C. Phillips (eds.), *Evaluation and Education: At Quarter Century*. Chicago: University of Chicago, 1991.
- Taylor, P. W. *Normative Discourse*. Englewood Cliffs, N.J.: Prentice Hall, 1961.
- Toulmin, S. E. *The Uses of Argument*. New York: Cambridge University, 1964.
- Toulmin, S. E. *Human Understanding*. Princeton, N.J.: Princeton University, 1972.
- Toulmin, S. E., Rieke, R., and Janik, A. *An Introduction to Reasoning*. New York: Macmillan, 1984.

DEBORAH M. FOURNIER is associate professor in the Department of Diagnostic Sciences and director of educational research and evaluation, Boston University.