Encyclopedia of Evaluation

Logic of Evaluation

The *logic* of a subject concerns such matters as its definition and the definitions of its major concepts, the nature of its relations with other subjects, the rules of inference that operate within it and in applying it, and logical disputes about these matters. What follows is a rather condensed treatment of some of these matters. It should be noted that these issues are not of merely academic interest because from them there follow many practical conclusions—for example, about what training is appropriate for evaluators, what kind of insurance coverage they need, how other disciplines can contribute to and can learn from them, and so on.

The Definition of Evaluation

Synthesizing the dictionary definitions of *evaluation* and *an evaluation* yields this: "determining the merit, worth, or significance of things; a report of such a determination." There is no need to deviate from the common meaning, and it avoids confusion in other people's minds to stay with it. Objections are often raised that such a definition excludes certain approaches to evaluation, such as utilization-focused evaluation, but of course it does not: They are simply approaches to this task, and the definition is strictly neutral about what approaches are best. Evaluators often feel that it would be more appropriate to define evaluation as what evaluators do, but that is a mistake. Evaluators do many things, such as conduct surveys and interviews and statistical analyses and, no doubt, going to church and watching television. The issue of definition requires us to focus on what they do that distinguishes them as evaluators, and the answer is just exactly their concern with determining value (i.e., merit, worth, or significance, depending on the context). Of course, evaluators do many other things that are not intrinsically evaluative, both on the way to an evaluative conclusion and in the course of business dealings that request these other services, for which many evaluators are extremely well qualified. If these other things were all that they did, an evaluator would be only a social scientist—a perfectly respectable, but somewhat limited profession. It is taking the extra step, from empirical or merely factual research to an evaluative conclusion that marks the evaluator as a practitioner working, at least partly, in a different discipline. That is the answer to the important question **[p. 236 ↓]** about the difference between evaluation and the usual kind of research in the social

sciences. Someone once said that the usual kind of empirical research is an attempt to answer the question, "What's so?" whereas the evaluator tries to answer the question, "So what?" One might add that the social scientist is often also concerned to find out "Why so?" and the policy analyst is often concerned with the question, "Now what?" Of course, for almost all of the history of the social sciences, social scientists' answer to the question "So what?" was simply that there could not be any scientific answer to it or, indeed, any rational answer. Evaluative questions were beyond the domain of science and reason, a mere matter of preference or taste. If that were true, there could be no legitimate field of evaluation.

The Validation of Evaluations

Which brings us face to face with the most central question in the logic of evaluation: How is it possible to justify answers to questions about value in a scientific or other disciplinary way? The skeptical answer to this question given by the social scientists —the doctrine of value-free science—did not come out of the blue. It was preceded by more than a century, beginning with Hume, in which the answer to this question was that it is logically impossible to do this; attempts to do it were said to have "jumped the is-ought gap" or, using Moore's later term, to have "committed the naturalistic fallacy." However, it has become clear that, like Hume's skepticism about the legitimacy of *causation*, his view, and Moore's, on this point, and hence that of the doctrine of value-free social science, was based on an oversimplified analysis of the logic of language. We can in fact infer validly from factual premises to evaluative conclusions by using definitions that bridge the gap: Because they are definitions, they do not count as value premises; that is, imported assumptions about values. The simplest cases are those in which propositions unpacking the meaning of "a good (or bad) X"—for example, "a good watch," combined with a number of facts about the performance of a particular watch— fully justify the conclusion that this is a good or bad watch (or X). This is what product evaluations do routinely, and it is logically impeccable. We commonly use exactly this kind of template in program evaluation and personnel evaluation, as well as in product evaluation.

There is a second way to justify evaluative conclusions, however—the barefaced way. Here we do use premises that express values and that are not definitionally true. They

are value premises, which can be directly validated in commonsense ways. There is no shortage of these: For example, premises expressing legal requirements or severe needs are often validated by legal research or systematic needs assessment and combine with facts about a program's performance to imply the merit (or lack of merit) of the program. Of course, they will be accepted only by those who regard adherence to the law or attention to needs as obligatory, but that includes the context of most program evaluations because they are done in a value-imbued context. Does this show that we have validated the evaluative conclusions? It does if we consider the assumptions of practical life to be sensible ones, and it does not involve making highly contested assumptions of an ethical or political kind. Of course, these assumptions lie outside the domain of the axioms of conventional social science, so accepting them marks another significant difference between evaluation and social science. It does not rule out the possibility that we might think it more sensible to expand conventional social science so that it takes on the support or refutation of such assumptions and could then claim that evaluation was part of social science. Many evaluators would favor that alternative, but until it is generally accepted, we must concede that this second method of validating evaluative conclusions marks a difference between the two disciplines.

However, the first approach to jumping the is-ought or facts-values gap (they are not quite the same, but both illustrate the same logical point) shows that we can bypass the great logical obstacle that was held to render our field inaccessible to scientific validation. How did this obstacle impede progress for so long?

Types of Value Claim

One of the sources of the fallacious belief that science was (or should be) value free was an oversimplified paradigm of value claim. It was generally taken that the prototypical example was exemplified in a claim like, "I value loyalty over industriousness in my employees" or, more simply, "I like white wines in general much more than red ones." This is, of course, a mere statement of personal preference and as such **[p. 237 ↓]** cannot be validated as having any wider applicability. Even if it asserts that "Italians value red wines well above white ones," it will still be merely a local truth, and a truth about a matter that is in one sense subjective. Still, one needs to be clear that such claims are perfectly verifiable by observation and evidence of other

kinds, such as buying habits, conversation, and so on, and hence are not subjective in the pejorative sense that puts them outside the domain of scientific testability. For many of the enemies of evaluation as a discipline—Ayer, Stevenson, and others— that was too much of a concession, and they argued that one should not treat these claims as proper claims at all; that is, as propositions or statements of a factual kind. They tried to subsume them under mere noises, but the effort to do so was implausible then and absurd now that we are less obsessed by the threat of legitimating evaluative statements.

In any case, there are other types of value claim that need to be distinguished from this one. For example, there are market value claims, such as a claim about the value of a house for sale in a particular location, and these have a standard method of verification recognized by the law and common sense; so these, again, are entirely testable and much less "subjective" because they do not have any particular connection to private states of the mind.

Then there are what might be called contextual value claims—claims that are prima facie factual but in a certain context refer to properties that are highly valued in that context and hence carry the import of an assertion of value. In the context of recruiting basket-ball players, for example, statements about an athlete's height are value imbued.

Finally, there are what we might call essentially evaluative claims, claims about the brilliance of Einstein or Chopin or Leonardo or about the excellence of the workmanship in a piece of Faberge jewelry or a great cathedral or the superiority of highly interactive modes of instruction over the lecture format or the disappointing results from the most heavily supported and widely implemented models of school reform or addiction reduction of recent years. These are what the fight is all about, and they are very unlike assertions of personal preference. Although in some cases they lack the support that entitles them to be called provable, in many other cases (e.g., in the claim about Einstein), they have more than enough support to justify the view that they are as well supported as the usual kind of particular or general scientific claim. These are the kinds of claim that evaluators aim to support by their work, and the suggestion that they are any less objective than claims about the advantages of one engineering design over another for certain purposes is no longer plausible—especially as that claim from engineers is simply one example of an evaluative claim, a type that is found

SAGE knowledge

throughout every scientific or other discipline and is known as an intradisciplinary evaluation. Throughout the long fight over the legitimacy of evaluation as a discipline, it was conveniently ignored that no discipline exists without standards of merit for the methodological entities with which it deals—the data, experiments, hypotheses, and work of its professionals—the application of which, of course, is simply one branch of the supposedly "forbidden" subject.

Branches of Evaluation

There are a dozen named branches of evaluation, of which the Big Seven are sometimes said to be the evaluation of products, performances, personnel, programs, policies, proposals, and portfolios. Some of these were practices half a million years ago and became skilled practices with specialist practitioners thousands of years ago, guilds after that, and then professions that blithely ignored the alleged impossibility of their subject across a century or more. Two fields of special importance and late emergence into recognition, although long practiced, are intradisciplinary evaluation and *metaevaluation* (the evaluation of evaluations). There are also many pseudoevaluative activities, thought by their practitioners to be branches of a legitimate discipline and often of considerable history, in which all the language and attitudes of evaluation are present but the basic requirements of testability and validated standards do not exist. Most (but not all) of literary criticism, art criticism, and wine connoisseurship clearly exemplify this syndrome. There are also subareas with great achievements that are still making many mistakes because they have not awakened to the fact that evaluation has evolved independently of their subarea, and they have lessons to learn from it. Consumer product evaluation is an example: The Consumers Union in the United States does great things but makes many mistakes because it thinks statistics and engineering alone are all it needs to design and do its work. The road-testing magazines are in the same category.

[p. 238 ↓]

Applied Fields or Areas of Evaluation

Most of the branches of evaluation can be applied to areas within each of the traditional subject matter areas. For example, within the fields of public health or social work or physics, there is plenty of program, performance, and personnel evaluation; some product evaluation and policy evaluation; and, of course, intradisciplinary and metaevaluation. Although there is not the slightest advantage to doing more evaluation if it is not good evaluation, in most areas, the commitment to evaluation is too small, both in resources and attitude, to achieve the maximum benefits. This is largely because of deficiencies in the training of professionals about evaluation; they all too often pick up a sense of defensiveness about it or a sense that it is something one has to do to get funded. The correct attitude is simpler: It is a minimum requirement of professionalism to ensure regular skilled external evaluation of one's work. Evaluation is a *professional imperative*, an obligation that is not only part of valuing good work in itself but a part of discharging one's obligation to society to do good work. Now that evaluation as a discipline has developed considerably, this obligation entails at least some effort to keep up with what is going on in the field itself, not just in already-implemented applications to one's own area. What has emerged in evaluation that might be useful to busy practitioners in applied areas?

The Fundamental Operations in Evaluation

In many applied areas, there is continuing carelessness and consequent errors in distinguishing among the core evaluation operations. Even in areas that pride themselves on precision and scientific method, such as measurement and testing, these errors are quite common. There are essentially five fundamental operations, with a couple of minor variations. The basic operations are grading, ranking, scoring, apportioning, and synthesis: All of these are familiar terms and practices, but it is extremely rare to find a practitioner or text that can define them correctly or distinguish them precisely. Quite different designs are required for each of these, and major failures to provide the help that is needed result from failing to distinguish them correctly. A horse race is a great approach to ranking, but not to grading, which will also require

scoring—and, in fact, measuring (a particular approach to scoring)—in this case, by using a stopwatch. If you have ever seen a scale in a survey in which the anchors are something like excellent, very good, average, below average, poor, or unacceptable, you have seen a design that confuses grading with ranking, which will produce uninterpretable results.

Two additions to the basic list are of great practical value: Profiling is the exhibition of grades (not scores) in several dimensions in bar-graph format: It is extremely useful in that it avoids the difficult and disputable task of synthesis in many cases where it makes it possible to do ranking without synthesis. Gap ranking is a refinement of ranking in which qualitative estimates of the intervals are added. It is standard practice in horse racing: "The winner is X by a head in front of Y; a length back is Z by a nose over W." It is extremely useful in a common context where ranking is required; for example, in the award of funds or prizes, where splitting the pot is possible.

Other topics in the logic of evaluation that may help indicate its utility today include the distinction between roles and goal of evaluation, where the *formative-summative* classification comes in: the function of program goals in program evaluation; the relation of explanations and recommendations to evaluations, which brings up the legitimacy and necessity of *theory-based evaluations* and evaluation-based recommendations; the distinction between criteria and indicators and how this shows that personnel evaluation cannot use style variables; and ethics as a branch of evaluation.

MichaelScriven

http://dx.doi.org/10.4135/9781412950558.n322
Further Readings

Scriven, M. (1991) Evaluation thesaurus (4th ed.) . Newbury Park, CA: Sage.

Scriven, M. Product evaluation—The state of the art Evaluation Practice 15 (1) 45–62 (1994) http://dx.doi.org/10.1016/0886-1633%2894%2990059-0