evaluation models. Ref: *Evaluability Assessment,* M. F. Smith, (Kluwer, 1989).

EVALUAND A generic term for whatever is being evaluated—person, performance, program, proposal, product, possibility, and so on—by analogy with "multiplicand", "analysand", and so on. If it is a person, the term "evaluee" is used here; although "evaluatee" parallels "evaluator", the analogy with "examinee" and "testee" and the greater brevity seemed more appealing. (Precedent exists for contracting the predicate term, for example, in "progenitor/progeny".) It is often possible and always desirable to avoid using this neologism, for example, by using 'candidate' or 'entry' or 'option'; but there are some cases in discussing the logic of evaluation where existing terms have inappropriate connotations, sometimes because the former suggests that people are involved, and the latter that some kind of competition is involved.

EVALUATE, EVALUATION Four possibly different senses of the term are distinguished here.

1. The key sense of the term "evaluation" refers to the process of determining the merit, worth, or value of something, or the product of that process. Terms used to refer to this process or part of it include: appraise, analyze, assess, critique, examine, grade, inspect, judge, rate, rank, review, study, test. A longer list, involving nouns as well as verbs, and including a number of terms that are only used evaluatively in special contexts, would also include: accredit(ate), adjudicate, allocate, apportion, appraise, appreciation, audit, benchmark, beta-test, check, check-up, classify, comment, criticism, determination, distribution, estimate, finding, field test, follow-up, gauge, interpretation, investigation, judge, mark, measure, monitor, overview, quality control, perspective, rank, referee, report, 'road test' or 'test drive' (now used metaphorically), scale, score, scrutiny, sea trial, survey, synthesis, tryout, weigh, verdict. The evaluation process normally involves some identification of relevant standards of merit, worth, or value; some investigation of the performance of evaluands on these standards; and some integration or synthesis of the results to achieve an overall evaluation or a set of associated evaluations. It contrasts with the measurement process, which also involves the comparison of observations against standards, in that (i) measurement is characteristically not concerned with merit, only with 'purely descriptive' properties, and (ii) those properties are characteristically unidimensional, which avoids the need for the integrating step. The integration process is sometimes judgmental, sometimes the result of complex calculation, very commonly a hybrid of the two.

In this sense evaluation is what distinguishes food from garbage,

lies from truth, and science from superstition. In short, it is the sine qua non of intelligent thought and action and in particular of professional practice. But it has also been an intellectual outcast for most of the history of intellectual investigations: the only one of the cognitive processes not to be covered in the science curriculum, the only one that so tainted articles submitted to professional journals that until the late 1960s they were automatically rejected.

Now, evaluation is not so difficult that one can explain its neglect as simply due to being 'put in the Too Hard basket', as the Australians say; it was in fact extensively practiced by those who denied its legitimacy. The explanation appears to be in part that for many people and organizations, evaluation is one of the most threatening phenomena in their experience. Some of them—the **valuephobes**—will lie, cheat, steal, and plot to avoid its occurrence or its impact, a phenomenon that often takes novice evaluators by surprise when they become the victim of character assassination. The *conscientious* practice of evaluation is thus more hazardous as well as more far-reaching than most applied social science research. People are often surprised to learn that Consumers Union, the bastion of product evaluation, was put on the Attorney General's list of subversive organizations in the war against Japan and Germany, and that a the current Director of the National Bureau of Standards was dismissed for providing, at Congressional request, an unfavorable although valid evaluation of a battery additive. They should remember that a large number of conscientious professionals in medicine as in journalism have lost their jobs for doing nothing more than what ethics requires with the results of good evaluations. Moreover, they should understand that the practice of evaluation is hard on evaluators for its own reasons, independently of the machinations of hostile evaluees. It is hard to maintain objectivity in the face of caused pain or joy and to decline bribe and threat combinations of various degrees of severity. The avoidance of evaluation thus achieves considerable support from many of those who would be obliged to do it, as well as those who would be subject to it.

If evaluation causes anxiety and the erection of defenses in many people, it is a source of power—over those who have not come to terms with it. As usual, this leads to efforts to reserve the power for a priesthood. This perspective on evaluation has an ancient history. In the Garden of Eden it is significantly the fruit of the tree of knowledge of good and evil that is taboo, indeed seriously taboo: God says, "in the day that thou eatest thereof, thou shalt surely die". The serpent inaugurates independent evaluation with this comment: "Ye shall not surely die: for God doth know that in the day that ye eat thereof, your

eyes shall be opened, and ye shall be as gods, knowing good and evil". The serpent is right on both counts. Nobody dies for the crime (an early counterexample to claims that God is omniscient and trustworthy), and God confirms: "Behold, the man is become as one of us, to know good and evil" (no mention of Eve, of course). The serpent turns out to be the only one of the four actors without moral flaw in the scenario—God lies, breaches contract, and acts unjustly, the other two try to blame someone else for their disobedience—and for this God curses the serpent "above every beast of the field". The parable thus tells us something about the risks of evaluation as well as about the connection of power to evaluative knowledge.

Myths apart, evaluation often acquires power because of its ties to possible action by decision-makers but more generally because of its potential threat to self-esteem. See **Balance of power, Whistle-blowers, Going native, Phenomenology of evaluation.**

2. The name of an autonomous discipline (now with its own Library of Congress classification); it refers to the study and application of procedures for doing objective and systematic evaluation (in the first sense). Semiautonomous applied areas include program, product, personnel, performance, proposal, and policy evaluation ('the Big Six'); other autonomous applied fields include **technology assessment**, medical or psychological evaluation, and **quality control**; other applications that reside within disciplines include **curriculum, sensory, aesthetic,** and **proposal** evaluation, and **literary criticism.**

Evaluation as a particular kind of investigative discipline is distinguished from, for example, traditional empirical research in the social sciences or from literary criticism, criminalistics, or investigative reporting, partly by its extraordinary multidisciplinarity. It typically requires consideration of costs, comparisons, needs, and ethics; political, psychological, legal, and presentational dimensions; the design of outcome studies; sources of bias; reactive effects; and a focus on the techniques for supporting and integrating value judgments—rather than on purely aesthetic matters, or on hypothesis-testing, theory-building, and taxonomy. Although aspects of the relevant part of these concerns—often quite primitive—are to be found in the social sciences, evaluation is not, contrary to the authors of most leading texts and references, a branch of applied social science, nor a study of human interventions, nor a subject whose intellectual origins are in the social sciences. It is a much older and more general discipline. Not only do systematic approaches to product and personnel evaluation predate the whole of social science by millenia, but so do the intellectual roots of the core discipline, the study of its methodology and models. Even if thought of in the limited sense in which it covers

some of the territory that social science should have been covering since it began, it still antedates social science, and has frequently used quite distinctive methods, not just existing social science tools. Examples include the key elements of functional analysis and its crucial connection to evaluation which came from Aristotle, the repertoire of logical assessment which goes to the preSocratics or earlier. Other elements like ethics, the study of self-referent evaluation, the relation of evaluation to political power, statistics, large-scale testing, cost analysis, models of legal reasoning, experimental design, and the logic of evaluative inference, all came from outside social science and often go back millenia instead of a century. Evaluation is properly conceived of as a discipline in its own right, an analytical discipline like mathematics (on the one hand, less precise, but on the other, much more general, useful, and fundamental to the human condition), covering a range of activities from quality control in manufacturing to the marking of student papers. Its occurrences in the social sciences should be seen as applications of the general discipline, not as applications of social science methods. It is one of the **transdisciplines**, although more multidisciplinary than any of them.

Evaluation—properly done—can be said to be 'a science' in a loose sense, as can, for example, teaching; but it can with equal justice to be said to be an art, an interpersonal skill like arbitration, and the logic implicit in the reasoning of judges and juries and literary critics and real estate assessors and jewelry appraisers—and thus not "one of the sciences". (See **Meta-evaluation**.)

Evaluation is normally contrasted with description, but this is only true in a particular context or from a certain point of view. "How would *you* describe the candidate, since you've known him for a long time?" is often followed by an account that is partly evaluative, and the questioner will not feel this is inappropriate. The function and hence the logic of evaluation is often to provide an extremely concise description of one aspect of something—its merit or worth. The letter grade that sums up a semester's work by a student describes the quality of that work. Indicator research is aimed at concise description of the state of the economy or of national health, and the indicators to be useful have to be evaluative, albeit sometimes **contextually evaluative**. A recent article on school process indicators makes clear that these would mainly be evaluative indicators (Andrew Porter, "Creating a System of School Process Indicators" in ***Educational Evaluation and Policy Analysis***, Spring 1991).

Many attempts have been made to distinguish (program) evaluation from research—typically, other social science research—for example, in terms of generality or generalizability, replicability, and

data types. It is true that the typical efforts of a contract evaluator or of someone whose job title defines them as an evaluator are more likely to be particularistic rather than general, by comparison with the typical efforts of a researcher. But this only corresponds to the difference between research chemists and those with the same qualifications who spend most of their time analyzing water samples for the water company. The latter are employed as practical or applied chemists rather than researchers, but both are chemists. Similarly, the applied evaluator does not own the domain of evaluation; the evaluation researcher, like the research chemist, is just as much a professional in the discipline. The slight difference in the way the term "evaluation" works, by contrast with the names of traditional disciplines, only means that one is a little less likely to say that the evaluation researcher is 'an evaluator' (whereas an inhabitant of either role is said to be 'a chemist')—but this kind of distinction is not unknown in the transdisciplines: the ethicist is someone working at the applied end, while the researcher working on metaethics, is called a philosopher. The distinction appears more attractive in the case of evaluation because there is no clearly identified location or name for—or tradition of—evaluation research in the academy. But eventually, "evaluation researcher" should be as recognizable a label as "cryogenics researcher".

In any applied field, meaning one that services clients with real-world problems, there is always one criterion for good work that is absent from the research field, namely the immediate utility of the conclusions. This will include their timeliness and cost-effectiveness, and bringing in these considerations means that the mission of getting the right answer will sometimes have to be compromised by conversion into 'getting the best answer possible under certain time/budget constraints'. In the end, this is not an alien approach, because most of what we refer to as 'laws of nature' are simply convenient approximations, but it does have to be clearly understood.

Stressing the difference between research and evaluation by using that phrasing is unfortunate, because it tends to support the same kind of mistake teachers make when they distinguish teaching from testing. The fact is that testing is an essential part of teaching; similarly, practical evaluation is an essential part of evaluation research, and research is an essential part of practical evaluation. Stressing this, for example by stressing that an estimate of generalizability (external validity) should be part of every program evaluation, is more constructive and productive than stressing the difference.

What distinguishes evaluation from other applied research is at

most that it leads to evaluative conclusions, and to get to them requires identifying standards and performance data, and the integration of the two. See **Formative and Summative evaluation.**

3. The term "evaluation" is sometimes, and unfortunately, used more narrowly to mean only the work done by professional evaluators. For example, a scientist on the National Science Board, when asked why NSF didn't evaluate its own evaluation procedures (e.g., by looking at such matters as interjudge agreement) said, "I don't think we can afford to do an evaluation of our procedures; it would simply divert sorely needed funding away from worthwhile proposals." Of course, she was really saying that she had informally evaluated the procedures and judged them sound enough. She was also recording some scepticism that the cost of a professional evaluation would pay off. But the question to which she responded was raised because of the disquieting evidence that the NSF proposal evaluation process is seriously flawed. (This evidence came from intelligent program administrators within NSF who have had some proposal evaluations done by two independent panels who were not informed that their work was being replicated.) The general idea in science, we are often told, is not to rely on assumptions and informal judgments but to do systematic study. However, she wasn't willing to apply this to her own foundation's procedures. Evaluation begins at home, and if you are evaluating proposals for the use of hundreds of millions of tax dollars, you should look carefully at how you do it, that is, evaluate the process, or get someone less ego-involved to evaluate it, because even a small improvement will pay off handsomely. Of course, the prospect of having these flaws documented involves the risk of loss of credibility and hence funding, so it lacks intrinsic charm.

In this criticism of her remarks no assumption is made that a single professional evaluator would in fact have done something useful. Although evaluation in the broad sense is a necessity for rational behavior or thought, and is indeed the only intellectual process common to all types of science, 'professional'—i.e., paid—evaluation is sometimes worthless, a sham, and/or excessively expensive (as is much other management consulting). Only a team of the best evaluators, working closely with program officers, experienced panelists, and NSF board members, could produce worthwhile results. But, given the facts mentioned here and similar findings elsewhere, those results would cover their costs many times as well as improving the justice of the procedures. Panel rating, as normally done, is a primitive procedure. See **Self-reference, Two-tier, Wild card, General positive bias, Cost-free evaluation.**

4. In what is normally thought to be a completely different sense, the term "evaluate" is also used in mathematics to mean "calculate the value of an expression"—for example, of a polynomial. The gap between these uses is not unbridgeable, however, as one notices when examining the term "**evaluation function**" in expert system work or "evaluative metering" in camera design. Each of these refers to the process of calculating the sum of several weighted values, just as in evaluating a polynomial. This is the key logical process which distinguishes evaluation from measurement—for example in cost-benefit analysis, or in giving a term grade to a student based on lab work, field work, attendance, and so forth.

EVALUATION ANXIETY Anxiety provoked by the prospect, imagined possibility, or occurrence of an evaluation. In the clinical context, this includes much social anxiety, test anxiety, pregame anxiety, and so on, and is a cause for concern only when it produces incapacitating affect or dysfunction. In the evaluation context, it is often something that deserves serious and direct attention, and dealing with it—especially the phobic version, **valuephobia**—calls for special skills and knowledge. Ref: ***Handbook of Social and Evaluation Anxiety,*** edited by Harold Leitenberg (Plenum, 1990). See also **Reactive evaluation.**

EVALUATION CONSULTANT See **Consultant.**

EVALUATION EDUCATION Consumer education (e.g., in home economics courses in the schools, or in the usual media presentations) is still very weak on training in evaluation, which should be its most important component. More commonly it simply involves doing evaluation on some limited topic with limited generalizability beyond the sources used. There are many other contexts besides those in which one's role is that of the consumer where evaluation education would be most valuable, notably in the manager role, parenting role, or the service-provider/professional role. Few teachers (or, for that matter, other professionals) have any idea how they or others can or should evaluate their own work or that of others, although this is surely a minimum requirement of **professionalism.** The last decades have seen considerable federal and state effort to provide reasonable standards of quality that will protect the consumer in a number of areas, but these agencies have not yet really understood that the superimposition of standards is a poor substitute for understanding the justification for them and having the skills to generalize them to new areas. **Evaluation training** is the training of (mainly professional) evaluators; evaluation education is the training of the citizenry in evaluation techniques, traps, and resource finding and is the only satisfactory long-run approach to improving the quality of our lives without extraordinary wastage of resources. It should begin with **critical think-**