THE NATURE OF EVALUATION

Michael Scriven
Claremont Graduate University

DEFINITIONS

1. Evaluation is the process of determining merit, worth, or significance (abbreviated here as m/w/s). In many but not all contexts these three properties are roughly the same as, respectively: quality, value, and importance.

2. An evaluation is the product of that process.

3. An evaluand is whatever is evaluated. (The term evaluee is often used for human evaluands.)

REFINEMENTS

The process of evaluation may be judgmental or inferential; either can often be done at the everyday/commonsense level or at a professional level. Examples: Judges of Olympic diving or skating, and referees in competitive sports, are examples of professional evaluators largely using judgment rather than inference; undergraduates checking an alleged new proof of Fermat's Last Theorem are amateurs using inference. English teachers in secondary schools, grading essays, include many that use one or the other approach exclusively; often the judgmental graders are superior. (Reference:…)

Professional level evaluation is distinguished by the use of validated skills or techniques that require extensive training, and by respect for standards of quality and ethics in the work; and it often involves an extensive commitment of time and other resources that is usually only possible for those paid to do this work.

The transdisciplinary view of evaluation

Definition 4. A discipline is: (i) a body of knowledge or skills with a common subject matter or field of activity, that is (ii) distinguishable from that of other disciplines, and (iii) often has some distinctive kind or kinds of investigative methods that are explicitly studied and debated (although they may be implicit rather than explicit in the common practice and teaching of the physical disciplines), (iv) that always include respect for logical requirements on evidence and inference (e.g., consistency, transitivity, defensibility); and a discipline is (v) normally—i.e., except in its birthing process—regarded as a serious field of work for a significant number of highly trained personnel, whose pursuit is (vi) normally expected to have produced at least some results of notable social, intellectual, or personal value, as judged by those in other fields.

> Note 4.1: this set of considerations is offered as sufficient; some may not be necessary conditions.

Note 4.2: besides the academic disciplines, of which evaluation is here argued to be one, there are also many physical disciplines e.g., boxing, dance, tai-chi, marathon training, regimental marching, and the artistic disciplines such as playing the piano and portraiture. Evaluation applies equally to them (details later), although it is of course not one of them; cf. 'intradisciplinary evaluation,' below.

DEFINITION 5. A transdiscipline is a special class of discipline that meets two conditions: (i) it has the status of an autonomous discipline, meaning that there are people studying it and its methods and practices for their own sake; but also (ii) its methods and results provide important tools used in a significant number of other disciplines. Statistics is a good example of a transdiscipline; mathematics is a more general example, since it includes statistics; logic is a supreme transdiscipline since its standards are part of the definition of discipline itself; and evaluation, so it will be argued here, is the only other universally essential transdiscipline.

The transdisciplinary view of evaluation is defined by a set of propositions dealing with various aspects of the view, as follows (for the most part these refer to professional evaluation):

Group A: Propositions referring to the geography of knowledge

Proposition A1. Evaluation can now be regarded as a discipline since it has: (i) a defined and fairly distinctive territory—the study of the process of determining m/w/s; (ii) some distinctive methods, described below; (iii) adherence to the usual research standards of logic and evidence accepted by the general body of disciplines; (iv) a work force numbering in the six figures worldwide, along with dedicated books, encyclopedias, journals, websites, professional associations and conventions; (v) contributions of some intellectually and socially valuable results, as judged by those not committed to it, several of which are set out in this work.

Proposition A2: Evaluation is a megadiscipline. Evaluation has a vast range of established sub-specialties (a.k.a. sub-disciplines) and an unlimited range of fields in which one or more of these can be applied professionally. Evaluation activity can be represented as a three-dimensional space, with axes for: (i) the sub-fields of evaluation, (ii) the application areas for evaluation, and (iii) methods of investigation. (There is some overlap between these.)

Proposition A3: Evaluation is a transdiscipline. It is not only a discipline, as argued above, but it is applied in every discipline as the part of applying logic and reason that relates to and indeed provides the entire coverage of meeting the standards of evidence and inference that justify claims about the quality of all research or knowledge in the discipline, and hence its claim to be a discipline at all. The reason astrology is not a discipline but merely a subject of study is simply a failure of the quality of its data and hypotheses, i.e., a failure to meet the quality standards of a discipline—and quality standards are the business of evaluation to define and supervise, at the general level.

[Fig 1 about here.]

2

The definition of the term 'transdiscipline' requires that evaluation is not only a discipline in its own right, with its many sub-divisions and applied fields, including theoretical as well as practical research, but it serves an essential role as a tool for much research in other disciplines. In fact, the claim of every discipline to that very title depends on the extent to which it maintains good standards in the assessment of evidence quality, hypothesis rating, and continued production of significant rather than trivial results—all of these being examples of 'intradisciplinary evaluation'.

Intradisciplinary evaluation is best classified as professional judgmental evaluation, in most cases, there being only a few heuristics and no algorithms to make it a matter of demonstrably valid inference. Learning how to do intradisciplinary evaluation is an essential part of becoming competent as a researcher, or advanced practitioner, in any discipline. The quality control system in almost every discipline depends ultimately on peer review. Of course, 'peers' are experts in that field; unfortunately, they are usually not experts in evaluation, with the result that peer review, as currently practiced, turns out to be an extremely defective system. It can be greatly improved, and suggestions for doing this are discussed later in this work.

> Note A3.1: when evaluators identify their specialty, they typically mention both a specialty area of evaluation and an application area; e.g., '(evaluation of) leadership training in the military.'

> Note A3.2: most of the names of the sub-areas in evaluation make some reference to an 'application area', in the sense of a focus on one type of evaluand; but the term in quotes is here used to refer to types of context (knowledge area, job environment) in which the evaluand type is located. This distinction between two senses of 'application area' is not arbitrary; it reflects the fact that the toolkit of specialized techniques for good evaluation is more substantially different between evaluand types than between areas of knowledge where the evaluand of interest happens to be located. But there are still significant differences between areas, e.g., leadership training for the military calls for considerable knowledge of the military's needs, capacities, and constraints that differs from those factors in training business or college leaders.

[Unpacking Figure 1, above:]

(i) Evaluation specialties

Many but not all of these are also properly called sub-fields (a.k.a. divisions or su-evaluations) of evaluation by type of evaluand. There are three groups:

> (a) 7 recognized autonomous sub-divisions (a.k.a. sub-disciplines or 'specialties') in evaluation, e.g., personnel and product evaluation. Each has thousands or tens of thousands of professionally trained practitioners, publications, conventions--the paraphernalia. The full list is: program evaluation (which includes the evaluation of processes, projects, systems, organizations, services), products (including produce), performances, personnel, proposals and plans, policies, portfolios;

3

(b) 3 new sub-fields identified as a byproduct of the development of a discipline of evaluation—intradisciplinary evaluation, meta-evaluation, and public evaluation (we'll define and discuss these later);

(c) 6 classical disciplines that have always been largely devoted to the practice of evaluation (ethics, logic), and certainly involved in it to an essential degree (medicine, engineering, political science, architecture). These are not usually or properly called sub-divisions of evaluation because they have long since been recognized as independent disciplines, and include much that goes well beyond evaluation e.g., huge areas of purely explanatory knowledge. But parts of them are, logically speaking, exactly that.

(ii) Application areas (a.k.a. environments) for evaluation. The range of things, and aspects of things, that can be evaluated is unbounded. Everything that exists or can be conceived can be evaluated: even God is judged good, and even quarks have charm. This is one reason evaluation is not only a megadiscipline but possibly the largest one in terms of scope. But in practice the turf on which evaluations are done—the environment of the evaluation where the clients and audiences live, demand, and consider the evaluation reports, can be usefully divided into different types. In the real world business of professional evaluation, the two most common areas of application are:

a. The fields of most common human endeavor, e.g., manufacturing, services, education, health, military, mining, travel, agriculture, management, art, and athletics. Any practical application of these fields involves and requires evaluation.

b. The academic disciplines, each of which is built on evaluation of every tool they use and every result they produce; and whose practical applications are also subject to and involve evaluation.

(iii) Methods of investigation or presentation. Surveying all the specialties in evaluation, it becomes clear that almost every known 'software' research method (by contrast with research hardware such as telescopes and the Large Hadron Collider) is useful somewhere in evaluation. Many of these are qualitative and many are quantitative; using both in one study, also a common and often optimal practice, is called 'mixed method' evaluation. There are frequently many alternative methods that will yield a good evaluation. Models etc.

 [Further discussion of some aspects of the transdisciplinary model.]

Intradisciplinary evaluation. The definition of the term 'transdiscipline' requires that evaluation is not only a discipline in its own right, with its many sub-divisions and applied fields, including theoretical as well as practical research, but it serves an essential role as a tool for much research in other disciplines. In fact, the claim of every discipline to that very title depends on the extent to which it maintains good standards in the assessment of evidence quality, hypothesis rating, and continued production of significant rather than trivial results—all of these being examples of 'intradisciplinary evaluation'. It is best classified as *professional judgmental* evaluation, in most cases, there being a few heuristics but no algorithms to make it a matter of demonstrably valid inference.

Learning how to do intradisciplinary evaluation is an essential part of becoming competent as a researcher, or advanced practitioner, in any discipline. The quality control system for it in almost every discipline depends ultimately on peer review. Of course, 'peers' are experts in that field; unfortunately, they are usually not experts in evaluation, with the result that peer review, as currently practiced, turns out to be an extremely defective system. It can be greatly improved, and suggestions for doing this are discussed later in this work.

ONE STANDARD LIST OF THE DISCIPLINES

· 2 Humanities

- 2.1 History
- 2.2 Languages and linguistics
- 2.3 Literature
- 2.4 Performing arts
- 2.5 Philosophy
- 2.6 Religion
- 2.7 Visual arts

· 3 Social sciences

- 3.1 Anthropology
- 3.2 Archaeology
- 3.3 Area studies
- 3.4 Cultural studies and ethnic studies
- 3.5 Economics
- 3.6 Gender and sexuality studies
- 3.7 Geography
- 3.8 Political science
- 3.9 Psychology

- 3.10 Sociology
- 4 Natural sciences
  - 4.1 Life sciences
  - 4.2 Chemistry
  - 4.3 Earth sciences
  - 4.4 Physics
  - 4.5 Space sciences
- 5 Formal sciences
  - 5.1 Computer sciences
  - 5.2 Logic
  - 5.3 Mathematics
  - 5.4 Statistics
  - 5.5 Systems science
- 6 Professions and Applied sciences
  - 6.1 Agriculture
  - 6.2 Architecture and design
  - 6.3 Business
  - 6.4 Divinity
  - 6.5 Education
  - 6.6 Engineering
  - 6.7 Environmental studies and Forestry
  - 6.8 Family and consumer science
  - 6.9 Health sciences
  - 6.10 Human physical performance and recreation*
  - 6.11 Journalism, mass media and communication
  - 6.12 Law
  - 6.13 Library and museum studies
  - 6.14 Military sciences
  - 6.15 Public affairs
  - 6.16 Social work
  - 6.17 Transportation