

Articles should deal with topics applicable to the broad field of program evaluation. Articles may focus on evaluation methods, theory, practice, or findings. In all cases, implications for practicing evaluators should be clearly identified. Examples of contributions include, but are not limited to, reviews of new developments in evaluation, descriptions of a current evaluation study, critical reviews of some area of evaluation practice, and presentations of important new techniques. Manuscripts should follow APA format for references and style. Most submissions are 20–30 double-spaced typewritten pages in length; longer articles will also be published if their importance to AJE readers is judged to be high.

Evaluation: Future Tense

MICHAEL SCRIVEN

INTRODUCTION

There are always two competing pressures on any attempt to project the future: the pressure towards realism (What *really will* occur?) and the pressure towards idealism (What *really should* occur?). One of the reasons that high-probability predictions about what will really happen are harder to make than high plausibility recommendations is that the latter typically influence the former more than vice versa (except, perhaps, in the stock market). This makes the second task more fundamental, at least in one sense (it may be causally antecedent). Moreover, there's always something seductive about helping to create the future compared to merely projecting it, and recommendations, if persuasively argued, do have some capacity to create and not just describe the future. For these two reasons, I'll focus here on the ideal-future option. And I think there's a good chance of getting to that state, although not without some tense moments.

SERIOUS EVALUATION

An immediate corollary of this decision is that we are restricting the discussion to serious, systematic, skilled evaluation—the kind that *should* be done—and only commenting in

Michael Scriven • Department of Psychology, Claremont Graduate University, 123 E. Eighth Street, Claremont, CA, 91711-3955; Tel: (415) 663-1511; Fax: (415) 663-1913; E-mail: Scriven@aol.com.

American Journal of Evaluation, Vol. 22, No. 3, 2001, pp. 301–307. All rights of reproduction in any form reserved.
ISSN: 1098-2140 Copyright © 2002 by American Evaluation Association.

passing on the spurious attempts at it that one finds occupying most of the verbal space in areas such as criticism of contemporary art and music, and wine connoisseurship. In the same way, a discussion of the future of science does not include any reference to Christian Science or astrology.

Note that some evaluation consists only of skilled observation, including a substantial slice of medical evaluation of a patient's condition or motion and so forth (the presenting symptomatology), so one must not restrict evaluation to matters involving long-chain inferences. Merit in gems, skill in teaching or diving, and illness in patients are often directly visible, at least to the trained eye. Similarly, evaluation does not always involve judgment; it may only involve measurement against established standards—for example, in evaluating performances in the high jump and other track and field events. So evaluation is not, as it is often said to be, simply the domain of value-judgments.

THE PRESENT AND THE PAST

Before we can say much about the future of evaluation, we need to get a grip on where it is now. People who spend much of their working lives doing program evaluation tend to think of that as evaluation, but it is only a minute fragment of evaluation, even today. While the mistake here is mainly a gross overgeneralization from one's personal experience, there is also a contextual consideration that makes this error much more attractive. It is that program evaluation represents a large slice of what has been discussed and formalized about evaluation. But that will probably—and should certainly—change, so we should not allow it to be a dominant influence on our projections. Nevertheless, it is an important difference between evaluation now and evaluation in the past, that is, the past of half a century ago, and that fact makes the error understandable. To correct its influence, we need to begin with a few words about the geography of evaluation as it is today, which should create a reasonable perspective on how little has changed and how much remains to be changed, for the better, of course. Note that books about evaluation at mid-century were often entirely about the testing of students, so the fact that the same kind of books today are entirely about program evaluation represents a change, even if it involves a similar mistake.

There are already, and have been for centuries (in some cases, millennia), about 20 recognized (and often named) fields of apparently skilled evaluation, ranging from the entirely unsubstantiable evaluative claims in the part of art criticism dealing with contemporary flat art (paintings, montages, etc.) and oenological aesthetics (wine-tasting), to the highly professional criticism of engineering designs that makes up part of contemporary product evaluation. Other respectable, and in many cases equally venerable, divisions of evaluation include: personnel evaluation, policy studies, performance assessment (notably of students/warriors/athletes), restaurant reviewing, program evaluation, road testing of vehicles, environmental impact evaluation, investment portfolio critique, the critique of play at cards and other games of chance including war, normative (i.e., applied) ethics, practical logic (argument evaluation), and applied decision analysis. These make up the greater part of the explicitly and long recognized territory of evaluation. But we should also make mention of two fields of evaluation that have only recently been identified as such, although both are of great importance.

The first of these is meta-evaluation (the evaluation of evaluations), the supradiscipline that keeps evaluation honest and shows it to be self-referent. It is now quite a strong

sub-discipline and can be conducted in several ways. The second of the newly identified fields is intradisciplinary evaluation, that is, evaluation as it is carried out within disciplines, for example, when the historian renders a verdict on the quality of Napoleon's conduct of campaigns, or the scientist chooses a hygrometer from those available or rejects a data set as of unacceptable quality. Because the only claim that a discipline has to being a discipline rather than mere opinion (even if decked out with some of the apparatus of science, e.g., astrology), or an elaboration of taste (the situation in wine criticism, e.g.) amounts to the claim that the evaluation it does in the course of doing business—the evaluation of data and hypotheses and theories—is valid and replicable, this branch of evaluation is the basis for credentialing disciplines as such. A single example from a discipline that was in the vanguard of the value-free dogmatism (which held that evaluation had no legitimate place in science): in the 2001 edition of the American Psychological Association's *Publication Manual* (APA, 2001, pp. 5–6), there is a nice section entitled Evaluating Content, with seven bulleted criteria of merit to be applied. So there is no science without evaluation, because without evaluation one could not distinguish science from pseudo-science, let alone good science from poor science. And distinguishing good science from bad science (e.g., in work submitted for publication, graduation, course grades, or promotion, and in one's own work) is an essential part of being a competent scientist. So, the real situation is exactly the reverse of that maintained by the value-free doctrine: it is not true that science is value-free, but rather that science is in fact totally dependent on evaluation for even the use of its name.

Judgments of value, far from being mere elaborations of taste properly resident only in one's private life outside of science, are in fact highly objective elements in every scientific enterprise, every scientific publication, and every plausible version of the scientific method (or indeed of any kind of critical thinking outside the sciences, since critical thinking requires the evaluation of arguments, assumptions, etc.).

EVALUATION IN THE NON-ACADEMIC WORLD

In addition to the role of evaluation in the disciplines, we should in this essay also put some emphasis on the vast region of everyday discourse and activity that is or involves careful evaluation. When we go shopping, many of us quite carefully evaluate fruit or meat before we buy it; when road work requires a detour, we select an alternate route to the store or to work; in crossing a street, we evaluate traffic conditions as hazardous or not; if we watch television or listen to a radio, we evaluate the offerings—the advertisements as well as the program content—before clicking the remote control; and we consider alternative jobs, careers, schools, living quarters, doctors, medications, vacations, exercise, diet, and lifestyles using evaluation of a casual or a thoughtful and complex if sometimes amateurish kind. But the latter cases are often the subject of “serious, systematic, evaluation” and the earlier ones often benefit from a grounding in it. Growing up and then living in that ocean of evaluation, we acquire substantial basic skills at these everyday types of evaluation, with the help of parents, siblings, and peers, and without them we would not survive for long. However, in the process we become so familiar with this “evaluation as an essential part of all practical life” that it ceases to be present to our consciousness as a distinctive and essential type of thought. Hence, even intelligent people swallowed the bizarre doctrine that evaluation was not a rational activity, was a mere expression of preference, and was certainly something to be excluded from science. In the future we may expect to see a more serious and systematic

approach to most of these everyday evaluation tasks, as we have lately seen in medicine with the revolution called "Evidence-Based Medicine."

EVALUATION AS A DISCIPLINE ITSELF

In recent years, we have made some progress towards clarifying the essential or core nature of evaluation. It has, in my view, now become clear that evaluation is itself a discipline, and a discipline of a familiar though not conventional kind. It is one of the elite group of transdisciplines, a term used in this context to refer to disciplines that are most notable for their service to other disciplines, although having their own autonomous status as well. These transdisciplines range from little but important ones like measurement, up through major ones like probability and statistics, key tools for most of the quantitative disciplines, to the all-encompassing ones—logic and evaluation.

To demonstrate its disciplinary status, one can list all its terms and relations, and spell out the logical principles governing them, which I have done elsewhere (Scriven, 1991). This process of clarification by unpacking the underlying assumptions and categories, here as in the case of measurement, statistics, probability, and logic, leads to the discovery of fallacies in procedures that had previously been acceptable. We find errors in the anchor points used in textbook examples of scaling, for example, where grading and ranking are often confused; we find errors in the standard procedures for the evaluation of teachers, for another example, where correlates of good teaching are confused with (i.e., used as) criteria of merit for teaching, an error that is both unscientific and unethical. So the enterprise of developing the core logic of evaluation pays off in practical endeavors as well as in the process of clarifying our thinking—and the fact that it does in itself contributes to the case for taking evaluation to be a discipline, for that categorization is to a considerable extent a pragmatic one. It's not worth identifying something as a discipline just because of some minor academic differentiation in subject matter; but if it not only has a large and readily separable territory (the study of merit, worth, and significance) and a distinctive logic (e.g., grading and ranking cannot be reduced to empirical measurement concepts), but also shows that its application leads to significant improvements in the standard disciplines, then it has (simplifying somewhat, because there are some other arguments as well) won its place in the pantheon of disciplines. Although it's true that in an extended sense one can view evaluation as a branch of logic, this is also true of mathematics. It's pragmatically more sensible to treat it as *sui generis*, an autonomous and specialized transdiscipline, normally taught separately from logic and understandable without serious training in the discipline of logic. Recognizing its distinguished status has begun, and it is time to make some bets about how it will continue.

EVALUATION IN THE FUTURE

In the Halls of Academe

In the middle of the last century, statistics was just emerging as a legitimate but not yet autonomous discipline, still taught only within the mathematics department. Today, it is taught in specialized forms within several other departments such as psychology, education, and biology, as well as having its own department in some university settings. Certainly

there are plenty of professors of statistics, including some holding endowed chairs. The future role of evaluation should be similar: a much greater degree of autonomy is essential to break the hold of the “What’s so?” approach which considers, for example, program evaluation to be simply applied empirical social science, a study of the effects and concomitants of interventions. It must be seen as *also* essentially involving the scholarly treatment of the “So what?” component, dedicated to determining the value or significance of the effects or concomitants, which brings in many topics and techniques not covered at all, or only covered in an extremely confused way, in current introductory methods textbooks for the social sciences. These include skills in applied ethical analysis, which must be seen as a branch *of* evaluation as well as an essential component *in* many evaluation areas, for example, program and personnel evaluation.

Now, evaluation as a transdiscipline services not only the social sciences but also history, geography, computer science, education, business & management, accounting, and the older sciences such as physics and biology. Links that enable majors in those fields, especially graduate majors in those fields, to take short courses in the discipline of evaluation need to be forged; in time there may be colleges where the specialized application of evaluation to these other fields will lead to courses and perhaps even professorships in Historical Evaluation, the Theory and Practice of Evaluation in Engineering, Auditing and Evaluation, and so forth. But outside these applied fields, respect for evaluation based on an understanding of its true nature and its value to other fields means that there must, sooner or later, begin to be autonomous departments of and chairs in evaluation. Without that recognition of autonomy, the development of the core logic of evaluation is unlikely to eventuate, let alone the development of the essential new curriculum materials in evaluation for the undergraduate and precollege strata of education (see below), and for stand-alone courses in evaluation for professional development and graduate educational service courses.

In Undergraduate Education

A major change in the conception of the general education or liberal arts degree will need to be made to bring it up to speed with the emergence of this new discipline, which breaches many of the best-loved dikes between disciplines. Other long overdue and related changes can and should be coordinated with this one, for example, (1) the need to replace traditional logic with the new discipline of informal logic (in which we do at least have one chair already) and greatly expand its slot; (2) the need to break the mold of the hypothesis-testing model as the paradigm of research in the social sciences (the best program evaluations are as sophisticated as any work in the social sciences but they are not hypothesis-testing exercises nor are they mere exploratory work prior to real scientific work, viz., hypothesis-testing); and (3) the need to get modern applied ethics out of the “an option within the optional philosophy concentration” slot and into the general requirements of education for our world. The combination of these efforts paints a picture that should prove enticing to some independently minded college, and will bring them much glory in the history of education. Let’s hope such a change doesn’t take half a century to percolate into practice. Many people have been lamenting the lack of a values emphasis in education, and the combination of teaching the practical skills of evaluation with teaching the key skill of practical ethical analysis would go a long way towards meeting the legitimate version of that demand. Why, there’s even a place for “respect for elders” in the values repertoire that should be considered in most social program and policy evaluation.

In K–12 Education

Along with elements of decision theory and informal logic, it is perfectly possible and certainly useful to teach basic product and personnel evaluation in the upper elementary school grades. It is really important to combat the natural tendency towards relativism that accompanies the adolescent rebellion against traditional family religious commitment. Now, it is ethically improper as well as educationally doomed to suppose that *indoctrination* with the values of the founding fathers (or mothers) of the native country or ancestral religion is an appropriate defense against that rebellion. It is time to try for something deeper, namely to teach the *foundations* of evaluation and ethics and the *skills of critical evaluation* that are implicit in the sterling deeds of (most) folk heroes.

Outside the Academy

The future should be one in which evaluation is: (1) fully accepted amongst educated people as one of the key disciplines; and (2) fully used in rational discourse and policy decisions. This will have a huge effect on the sensible distribution of resources, and hence on improving the lives of those in need. But it will also have a huge influence on the evaluation of research and technology, continuing sources for increased wealth and investment and education. And the sleeper application of evaluation, intradisciplinary evaluation, when treated explicitly, will have a large effect on pure science from which often comes inspiration for the applied fields and for our minds. This has only just begun to emerge as one of the payoff areas for reconsidered evaluation. We have already seen ways in which the evaluation of proposals for research support needs to be radically reshaped, for example, to support more “out of paradigm” research. It has been suggested above that the logic of models for scientific research in the social sciences needs to be rethought. The pathetic history of the worship of statistical significance as a guide to scientific importance reminds us of another area where critical, fundamental, evaluation of scientific method has begun to change it, after much too long a time—more than 30 years after the first major attack was published by distinguished authors in book form (Denton, 1970).

IN CONCLUSION

The above examples are just some of many illustrations one could provide of the frightening extent to which the vaunted scientific method—that mighty engine of reason, when properly applied—is so poorly conceptualized *or* applied (it scarcely matters which) that science is prevented from achieving some important goals because of its own mistakes, and, one might add, its own overblown sense of its own success. Possibly the worst example of this is the way in which most of 20th-century science was wedded to the value-free doctrine while practicing evaluation with almost every stroke of pen or keyboard, and thus prevented from application to many of the most urgent needs of humanity. If no other benefit ever results from evaluation, we should never forget that one lesson in how the high and mighty make high and mighty mistakes.

Just as the vaunted source of democracy, the Greek city-state, was in fact a society built on *slavery*, and just as the authors of the Declaration of Independence did not think women and slaves deserved to have *their* independence recognized, so we need to watch most closely

those institutions which promise the most, including of course this paean to the virtues of evaluation. We do not cheapen our commitment to the scientific method, or democracy, or the values of this nation, by critical attention to the extent to which their high priests practice what they preach: we only instantiate it. It is those who support their own interpretations by saying, "Those who are not for us are with our enemies," that repudiate the essence of scientific method, democracy, and freedom—for critical evaluation is at the very heart of each.

REFERENCES

- American Psychological Association. (2001). *Publication manual of the American Psychological Association*, 5th edition. Washington, DC: Author.
- Denton, E. (1970). *The significance test controversy: A reader*. New York: Aldine de Gruyter.
- Scriven, M. (1991). *Evaluation thesaurus*, 4th edition. Thousand Oaks, CA: Sage.