jmde

http://www.jmde.com/ *Editorial*

Originally published on January 31, 2009
Revised on February 2, 2009

Meta-Evaluation Revisited

Michael Scriven
Claremont Graduate University

Next month will mark the passage of forty years since I published my first article about "meta-evaluation"[1] (Scriven, 1969), a term I had invented somewhat earlier in a report to the Urban Institute, who had asked me for help in dealing with the noncomparability of the evaluations they had commissioned for several housing projects. This year also marks the elevation of the concept to the status of an independent category in the latest edition of the Program Evaluation Standards (Joint Committee, 1994).[2] So the editors of *JMDE* thought it might be time to take another look at this process and its products, and, as promised, we have provided a special section in this issue on the topic. Many thanks for the much-appreciated contributions that were sent in for that section! Before reading them, and hence without implying any criticism of them for ignoring or disagreeing with what follows, I'm going to mention some reflections on the notion that now seem to me worth stressing. They—and the other articles in this issue—may inspire reactions from you: Please put them on paper or screen, and we'd be happy to consider them for the (nonrefereed) discussion section of a later issue of *JMDE*.

I introduce these points with the general thought that there seem to be three categories of question that commonly arise about meta-evaluation, namely inquiries as to (i) exactly what it is, (ii) how it can be justified, and (iii) when and how it should be used. In what

[1] I continue to hyphenate meta-evaluation, partly because for non-English speakers, there is some difficulty in knowing how to handle the adjacent vowels in metaevaluation, but practice varies on that.

[2] Footnotes on the history of the concept should perhaps include (i) the original reference, which is: Scriven, M. (1969), An introduction to meta-evaluation, *Educational Product Report,* 2, pp. 36–38; (ii) the need to avoid the mistake of many, including the World Bank thesaurus, of confusing meta-evaluation with meta-analysis. The latter is an ingenious statistical technique, introduced and christened by Gene Glass in the 1970s and applicable only to *a set* of *quantitative* studies (which may *or may not* be evaluative), for synthesizing their results *in terms of statistical significance*. Meta-evaluation, by contrast, *only* applies to evaluations; is *frequently* (and properly) applied to just one at a time; may assess an evaluand that is entirely *qualitative;* and refers to *any* process of evaluating it, or them, or the product of that process. Note also that meta-analysis of a set of evaluations begins with a limited meta-evaluation of each to see if it meets certain minimum standards of validity. Also note that meta-analysis of nonevaluative studies is often useful as part of an evaluation. (iii) Meta-evaluation may also be applied to the entire practice of evaluation or a subdivision of it. Here's an example from recent *New York Times* review of a book on architectural assessment: "Huxtable's work remains the gold standard of criticism—and not just the architectural variety—because she brings to the job a rare combination of aesthetic certitude and roving curiosity" (Davidson, 2008). (Of course, the general claim in this quote is not universally true, since—for example—much excellent product evaluation [think medicinal drugs] eschews an aesthetic component.)

follows, I say something about each topic—definition, justification, and application.

1. *Meta-evaluation is the consultant's version of peer review.* One sometimes encounters the objection to meta-evaluation that its use looks like a confession of incompetence. But almost every scholar is quite comfortable with the notion of peer review, i.e., with the idea that their submissions for publication will be sent out to competent, independent professionals for evaluative comment before appearing in print. That process is a valued (or at least accepted) part of the context of most scholarly work today. But it is not normally part of the context in which a consultant works—and most evaluators operate in that role. They do their work and then submit the results directly to the client or other audience. To get an expert's comments on that work before submitting it is therefore not a confession of weakness, but just a recognition that—as experts often say to their clients—an independent expert's look at one's work often generates insights that help to improve it.

   Accepting the importance of that truth is part of the culture of international scholarship today. But that was not always so. Einstein was very upset when he moved to America and discovered that the editor of a leading journal to whom he had submitted an article had actually sent it out for someone else to scrutinize without getting Einstein's permission. The reviewer had in fact criticized the paper, making a point that Einstein said, in complaining about this process to the editor (and withdrawing the article), was completely invalid. In preparing the paper for later publication elsewhere, it turned out that the reviewer had been right and Einstein wrong, so this story has an ending that provides a lesson to all of us: Even Einstein could benefit from peer review.[3] We can call this "the argument from peer review."

2. *Meta-evaluation is a demonstration that (at least some) evaluators practice what they preach.* We shouldn't make the usual pitch to clients that their work would probably benefit from serious (and especially external) evaluation, which is true enough, and then act in a way that ignores this advice. Putting the point in another way, if the client—or anyone else—asks, "Who evaluates the evaluator?" it's better to have an answer that concedes the point and names the role as a standard practice than to act as if the question doesn't apply to you. We might call this "the argument from self-reference"; qualitative researchers might call it "the argument from authenticity." Of course, there's always the question of whether the contract provides enough time and budget to cover the requirements for a meta-evaluation, but on projects of any significant, size the evaluator should make sure it does.

3. *In meta-evaluation, as in all forensic work, check the pulse before trimming the nails.* It's all too easy to respond first to the first thing you see that needs fixing. But all evaluation is essentially critical thinking in a systematic, skilled, and localized way, and the first lesson in critical thinking is to check out the basic assumptions before fiddling around with the finer points. It's easy for you, working as a meta-evaluator, to start at the beginning of an evaluation report and work

[3] Thanks to Chris Coryn for this story; it's in his doctoral dissertation (2007) and his book, *Models for Evaluating Scientific Research* (Coryn, 2008). Don't forget Einstein's implicit point, however: He was not just being a prima donna in this story. Peer review for traditional paper journals does involve some risk of pre-emptive plagiarism by the reviewer. However, that risk does not usually exist with meta-evaluation, since it's rarely done in a highly original *and generalizable* way. And it is a notable advantage of online publication that the compressed timeline makes plagiarism much less of a problem.

your way through it, picking up all the problems you find on the way. But that tends to take too long and can get you lost in the byways. You may in fact entirely miss the big problem, which is that the evaluator has focused on the wrong thing entirely—for example, the process instead of the outcome, or vice versa, or only one (or both) of these instead of both and more. It's not that you shouldn't read the whole report meticulously, it's that you need to do it while spending half your time thinking about what the report is *not* doing, what it *might* be doing, or what it *is doing viewed from a holistic point of view* rather than the piecemeal basis that a line-by-line approach prompts you towards. To do it correctly may mean you have to skip some of the report, but that's the right choice if you can't extend the timeline. This is the "argument for (scientific) fundamentalism."

A possible implication of this argument is that the attempt to list all the criteria of merit for a meta-evaluation in a task-specific checklist,[4], and use that as an outline, which I did as soon as anyone and have supported strongly ever since, might also be said to nudge the meta-evaluator towards a piecemeal approach. I think this is not a real problem, certainly not to the same degree as with the seriatim approach criticized above, but in either case it is extremely important to check for underlying assumptions. For example, in reviewing a Department of Defense evaluation of the conduct of the invasion of Iraq or Afghanistan, or Panama or Grenada, by U.S. forces, the meta-evaluator should expect the evaluation to consider the question of whether U.S. forces should be there at all, or at least expect it to justify the failure to consider that question, since for many stakeholders—including many whose lives were lost or at risk—this is the most important issue of all and is arguably a question of military strategy.

4. *A partial meta-evaluation is better than none.* In the real world, a complete meta-valuation—which would involve (i) checking or redoing the data-gathering and (ii) the values-validation[5] as well as (iii) the design, (iv) the data-analysis, and (v) the conclusion(s)—are rarely done. But a review of, for example, the second and third, or even just the fifth, of these components can be extremely valuable, even without the others. At the applied level, this is a more important point than may appear at first sight. For example, recent work has shown that trained evaluators can often spot serious weaknesses in highly localized evaluations without local knowledge by pointing out flaws in the logic of evaluation on which all evaluations depend (e.g., Coryn & Scriven, 2008). This suggests that the common practice of setting up review panels (e.g., of medical facilities or engineering colleges or business schools or businesses) consisting entirely of specialists in the discipline is flawed. In general, including one evaluator, even if she or he replaces one specialist, is good insurance against fatal design errors (including ingroup bias).

5. *Make the most of meta-evaluation: It has more than one function.* Not in the original article, but later, I suggested a practice that is sometimes worth considering (especially for credibility reasons)—namely having the meta-evaluator send his or her report simultaneously to you (the primary evaluator) and the primary client. Doing this (i) follows the analogy to improved forms of peer review, (ii) gives the client a sense of

[4] For example, the Meta-Evaluation Checklist (MEC) has seven checkpoints: validity, credibility, ethicality, utility, robustness, cost-effectiveness.

[5] For example, this might include redoing the needs assessment, which is often the most important source of relevant value, and very often the Achilles' heel of the evaluation.

getting something like a warranty to cover your use of their funds,[6] (iii) often provides an independent confirmation of your findings (or at least your methodology), and (iv) gives you a chance to respond immediately, in your final report, to criticisms that might surface later. But this move towards transparency is not an ethical necessity, since using a meta-evaluator is, in my view, good standard practice for an evaluator aiming to do the best work possible, and hence needs no special permission from or direct communication with the client (unless it would violate confidentiality constraints). I now think that a better practice, if time and patience allows, is to have a draft meta-evaluation based on your draft evaluation sent only to you, the evaluator, for reaction and possible modification of the draft evaluation. If the latter is then modified, it can then be next sent to the meta-evaluator for possible modification of the draft meta-evaluation. Then the final meta-evaluation is sent to the client and you, and you can submit your final evaluation report, which will include your response to the final meta-evaluation. There is one ethical imperative that must be retained: The last best effort by the meta-evaluator must go to the client intact.

6. *Choosing a meta-evaluator requires the same integrity that all evaluation requires.* In general, don't use friends—unless their hero is Gregory House (the ruthless diagnostician on the eponymous TV series). And don't just go for the "critical friend" that Fetterman recommends, though that's better. Go for a "smart enemy" or at least a tough rival. They are (i) more motivated to find fault, (ii) less likely to be biased by the modest fee to deliver a friendly report, and (iii) probably working from a truly different perspective, i.e., different basic assumptions. Just what you need![7] If you discuss meta-evaluation with the client, perhaps as a way to provide some warranty of quality, you might encourage her or him to choose the meta-evaluator. How would they know who to pick? A good first choice might be whoever came second in the competition for the evaluation contract, if you weren't sole-sourced. Or they may know other evaluators. Or (in the U.S.), they could call the current president of the AEA for suggestions (in other countries, the president of the nearest national association of evaluators, see *eval.org* for a list).

7. *Any systematic approach to evaluation—in other words, almost any kind of professional evaluation—automatically provides a systematic basis for meta-evaluation.* Obvious examples of such approaches are the checklist approaches, e.g., CIPP (Stufflebeam, 2007), the Program Evaluation Standards (PES) (Joint Committee, 1994), and the Key Evaluation Checklist (KEC) (Scriven, 2007). Each of these, whether explicitly or implicitly, involves a claim to comprehensiveness. Hence one can do a systematic meta-evaluation by checking the evaluation against one of these lists of what an evaluation should cover and how this coverage should be done. These approaches always provide a handy alternative to using one of the specific checklists designed for meta-evaluation, e.g., the MEC defined in a footnote above. It's always desirable to use a different one than is used by the primary evaluator, but even using the same one can yield very good results. However, as Jane

[6] Actually giving a warranty has long been discussed in the consulting world. It has great difficulties because it's hard to define what you're undertaking to provide in a testable way. For example, you might say you undertake to provide good advice, but you can only prove it's good advice if they implement it, which they won't do if they are in the mood to take you up on the warranty.

[7] If you use a "smart enemy" and plan to send the results to the client, as suggested in point 5 above, it may be wise to explain your reasoning for this choice to the client in advance.

Davidson reminds me, you're then pinning your effort to the validity of these checklists and even if they are intrinsically valid, they may be *applied ritualistically*, and thus miss important points.[8]

8. *Meta-evaluators need meta-meta-evaluators.* No, that doesn't lead to an infinite regress, partly because funding is finite and partly because the evaluator/meta-evaluator relation is reciprocally evaluative. The meta-evaluation series is quickly convergent. If the program costs $100K, the evaluation might cost $10K, the meta-evaluation $1K, and the meta-meta-evaluator (MME) (and partner) gets taken out to a good dinner. The MME can use any checklist or a holistic approach. The meta-evaluation checklists are, roughly speaking, set at one level of greater abstractness than the more practice-oriented KEC, CIPP, and PES checklists. Hence they do not need to list meta-evaluation as a component, but for all the reasons here, the meta-evaluator should try to use a meta-evaluator—that is, a critique by an independent evaluator—where possible. This is good advice and old advice in all scholarship—and in the practical arts, too. Which brings us to the next point…

9. *Fundamentally, meta-evaluation, like evaluation, is simply an extension of common sense—and that's perhaps the best defense against the suggestion that it's some kind of fancy academic frippery.* Here's a little evidence for that view: You have no doubt noticed the vast growth of "user reviews" of products that are sold online. Not only are these attached to product descriptions by many big retailers (e.g. Amazon, also including manufacturers such as Dell), but there are many Web sites devoted solely to these evaluations (e.g., Epinions, Church of the Customer, Yelp, TechRadar). This is one aspect of what I call "public evaluation" (i.e., evaluation in the public domain done by nonprofessional evaluators, such as politicians, industry leaders, consumer magazine reviewers, and [many] government officials), and it's one that particularly facilitates not only incompetence but corruption (e.g., the use of paid shills and employees to submit reviews). Naturally, some of the specialist sites have been struggling with the problem of detecting these flaws. That is, they are getting into meta-evaluation. A recent issue of *PCWorld* reviews these efforts in an article titled "Can You Trust Online User Reviews?" (Luhn, 2009). The suggestions there include both good and bad ideas, and it's clear that Amazon, Yelp, and Epinions use some excellent, admittedly imperfect, crosschecks, although several others mistakenly complain that nothing can be done. The approaches are, respectively, (i) public access to all other reviews done by every reviewer, (ii) use of ingenious computer-driven algorithms, and (iii) use of a user panel that interrogates reviewers. Clearly, (iii) is the best approach, even though it will deter many potential reviewers, so take Epinions' reviews more seriously. Once again, common sense in product evaluation leads to some good evaluation practices that may be transferable to other fields of evaluation.[9]

10. In general, looking at the number and qualifications of our other contributors on this topic, I conclude that meta-evaluation is alive and well—and getting better, as are the defenses for it. Since meta-evaluation includes all justification—as well as all

[8] She hopes to spell this out in more detail in a note in the discussion section of our next issue, where I also hope to put in some reactions to the other articles in this issue. And you, dear reader, should consider adding your own thoughts—about any of the contributions to this issue or the issue itself (e.g., should we do more special topic sections, on what topic, was this one useful, etc.)

[9] An earlier example was the use of goal-free evaluation, which is close to standard practice in much consumer product evaluation, e.g., road testing.

criticism—of evaluation itself, the status and possibly the survival of evaluation depend on it. It is perhaps fair to say that meta-evaluation is the conscience of evaluation, just as evaluation is the conscience of society. Long may both continue to improve!

References

Coryn, C. L. S. (2008). *Models for evaluating scientific research: A comparative analysis of national systems*. Saarbrücken, Germany: VDM Verlag.

Coryn, C. L. S., & Scriven, M. (Eds.). (2008). Reforming the evaluation of research. *New Directions for Evaluation, 118*.

Davidson, J. (2008, December 28). Critic on site [Review of the book *On architecture: Collected Reflections on a century of change*]. *New York Times*. Retrieved January 10, 2008, from http://www.nytimes.com

Luhn, R. (2009, January 1). Can you trust online user reviews? *PC World*.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage.

Scriven, M. (1969). An introduction to meta-evaluation. *Educational Products Report, 2*, 36-38.

Scriven, M. (2007). *Key evaluation checklist*. Retrieved January 22, 2009, from the Western Michigan University Web site: http://www.wmich.edu/evalctr/checklists/kec_feb07.pdf

Stufflebeam, D. L. (2007). *The CIPP evaluation model checklist*. Retrieved January 22, 2009, from http://www.wmich.edu/evalctr/checklists/cippchecklist_mar07.pdf