

# An Evaluation of Michael Scriven's "Minimalist Theory: The Least Theory that Practice Requires"

WILLIAM M. K. TROCHIM

## INTRODUCTION

In his Plenary Address entitled "Minimalist Theory: The Least Theory that Practice Requires," Scriven (1998) demonstrates the compelling quality of his thought that has led to his continuing influence on the theory and practice of evaluation. By any criterion checklist I can imagine, his work would be judged to have continuing merit. In my remarks on this address, I hope to summarize what I understand Scriven is saying and make a few evaluative judgments of my own along the way.

Scriven has three major points, and numerous other observations that I will not dwell upon:

- He argues for a "minimalist" use of theory in evaluation.
- He outlines a definition of evaluation.
- He argues for the value of checklist methodology in evaluation.

## Minimalist Theory

Scriven's first point is a direct challenge to the theory-driven view of evaluation. He distinguishes between what he calls "internal" and "external" theories. In program evaluation, internal theories describe the program itself—its components and operations. External theories describe how the program produces its effects or outcomes. To me, it sounds like Scriven's internal theory is essentially what social researchers mean by the construct of the program. And, his idea of external theory sounds like the social researcher's notion of a causal



William M.K. Trochim

---

William M.K. Trochim • Cornell University, Department of Human Services Studies, College of Human Ecology, Room N136C, MVR Hall, Ithaca, NY 14853; Tel: 607-255-0887.

---

American Journal of Evaluation, Vol. 19, No. 2, 1998, pp. 243-249.  
ISSN: 1098-2140

All rights of reproduction in any form reserved.  
Copyright © 1998 by American Evaluation Association

---

relationship. My reading is that he is reinventing the ideas of construct and internal validity that we are familiar with (although his "external" is the old "internal" and his "internal" is the old "external" validity, which is what construct validity was classified as). His key point about these two types of theories is that:

... it is quite possible to understand the external theory of a program, i.e., what kind of effects certain kinds of interventions can have, and know how to avoid some common side-effects, without knowing anything about how the program works internally to produce the outputs, i.e., about the internal theory of the program. (Scriven, 1998, p. 60)

What are we to make of this? He seems to be saying that we can have internal validity without having construct validity, that we can establish a cause-effect relationship without a detailed explication of the construct of the cause. Scriven seems here to take an instrumentalist approach that is wholly consistent with the black-box experimentalist view of research—it doesn't matter how the program works, only that we determine whether it does. This black-box mentality was the original jumping-off point for theory-driven evaluation, the very thing that people like Chen and Rossi (1983, 1987) were criticizing in their original statements.

It is undeniable that instrumentalism has its place. If we find something that works, we have found something of instrumental value, regardless of the theoretical reasons for why it works. If we find a cure for AIDS or figure out how to reduce drug dependency, why should we care how these interventions work? I have two problems with this view. First, it turns out to be terribly hard to find interventions that work. Doesn't theory have an important role to play in helping us decide what interventions to try? Second, if we do an outcome evaluation and determine whether or not an intervention works, doesn't it help us to improve that intervention or design alternatives we might try to know *why* that program may be working or not? These are the kinds of concerns that led to the rise of theory-driven evaluation in the first place.

In the literature on validity we have long supposed that there is an inherent tension between "internal" validity—the search for a cause-effect relationship, and "external" validity—including the explication of the constructs. Methods that enhanced one (such as experimental approaches for increasing internal validity) often seem to diminish or be in opposition to the other (such as detailed explication of the construct of the cause). I have argued elsewhere (Trochim, 1985) that this opposition may be overemphasized. Instead, I argue that greater emphasis on construct validity (one aspect of external validity) actually enhances our chances for determining internal validity. In more prosaic terms, if we do a better job of explaining what the program is, we have a better chance of determining whether it is working. Why? Because a more detailed explication of the construct—the program or intervention—allows us to state more precisely the pattern of outcomes we expect it to create. If we treat the program like an on-off switch, we have to distinguish its effects from all the other factors that could lead to an on-off result. This is why we emphasize the role of good control or comparison groups in experimental contexts. On the other hand, if we treat the program as a detailed fingerprint, it may be easier for us to distinguish its unique traces in the pattern of outcomes. This naturally gives rise to the idea that pattern matching has an important role to play in evaluation. But we can't do pattern matching well unless we understand enough about the program—the finger—to spot its distinctive effects—the print. And understanding the patterns involved in a program depends on having some idea of that program's com-

ponents and how it works. Furthermore, we aren't necessarily forced to choose between these two approaches. We can do both in an evaluation. In fact, my point is that we will have a better chance at seeing the causal effects of the program if we enhance our experimental endeavors with good detailed explication of the program construct.

How does this translate in practice? Imagine an experimental study of some educational program. An experimental outcome evaluation—essentially Scriven's external, black-box, instrumentalist study—would enable us to assess whether there is evidence that the program causes demonstrable effects on one or more indicators. But we seldom expect that a program will affect all outcomes equally. If we have a theory about how the program works, we are in a better position to state a priori that the program should affect outcome A more than B and that, in turn, more than C, and so on. Imagine that we have lots of outcome measures, say ten or even a hundred. In the experimental context we can determine, within the bounds of probability, whether the program affected A or B or C. But if we know more about how the program operates—that is, we have a more detailed program theory—we can additionally say how much each outcome might be affected. Suppose, for the sake of argument, that we run our statistical analysis and determine that none of the outcomes is significantly affected by the program. We have here the classic graduate student's dilemma—none of the findings in our dissertation study is significant. Further, suppose that we have a more detailed theory of the program that allows us to predict outcomes, to predict the ordering of effects. We might find that—even though no individual outcome is statistically significant—there is a strong significant relationship between the order of outcomes we would predict and those we obtain. This is the pattern matching hypothesis. It requires that we have a more detailed theory of the program and how it operates. It enables us to examine the unique trace of the program in the outcomes. We may find that our theoretical pattern is only partially correct—there are a few outcomes that don't conform. This can lead us to revise our understanding of how the program works, or to change how we measure the outcomes, or both. Doesn't this more theoretical approach extend the simple black-box model in ways that are useful and interpretable? There seems to me to be nothing contradictory between these approaches—the experimental instrumental one and the more theory-driven pattern matching one. We need to find better ways to incorporate them side-by-side in our evaluation studies. The intuition of the desperate graduate student and of the theory-driven evaluator is essentially the same—just because we haven't detected demonstrable outcomes doesn't mean the program is not working. If we know more about how the program works we may be able to see its effects more clearly. And sometimes we may miss program effects altogether if we don't have some idea of how the program works.

I hasten to add that I am not an apologist for theory-driven evaluation as currently espoused. But my problem is different from Scriven's. Unlike him, I am in favor of greater incorporation of theory into evaluation efforts. But my problem is that too many evaluators seem to emphasize the role of academic, social science theory. We seem to think that social scientists or management theorists have an inside track on developing sensible theories. My view is that the people in the best position to tell us about the theory of a program are the ones most familiar with it—the people who initiate it, develop it, implement it and receive it. The problem is that many of those people work in the practical world—they're not used to stating their implicit theories overtly in ways that we can formally evaluate. This is why I have spent

so much of my career trying to develop simple methods like concept mapping to help draw out implicit theories of different program stakeholders. I think we need to be examining program theory as an integral part of evaluation. But let's include the implicit theories of the people closest to the program and not deify the value that more academic theorizing may have.

I conclude this rambling critique of his minimalist approach with a few questions I would like to see Scriven address:

Are you saying that evaluation should deal primarily or exclusively with what you call "external" or causal theories—with determining whether interventions work—and not with how they work?

What is your view about the role theory plays in developing interventions? And, shouldn't evaluators use such theory—whether it's explicitly stated or implicitly held—as an integral part of our evaluation work?

In the end, I do not find Scriven's minimalist theory compelling. While he correctly points out that an evaluation does not require theory to be useful, he fails to see how theory can enhance our ability to identify whether a program works.

### **So, What is Evaluation Anyway?**

In his second major point, Scriven revisits the unending battle to define evaluation. I really struggled to try to understand what bothers me so much about his conclusions. He presents a view that is at the same time both simple and attractive. Listen to what he says:

All scientific work rests on a continual, repeatable, intersubjective, multiple processes [sic] of evaluation, such as the evaluation of data, hypotheses, classification, taxonomies, measurements, instruments, experimental designs, interpretations, theories, scientific papers, student work, scientists, and so on. (Scriven, 1998, pp. 65-66)

Or, consider this passage from another of his recent speeches:

If you open a scientific journal in a randomly chosen field, at a randomly chosen article, it is likely that you will find a presentation which begins with a review of previous work. This review is highly selective and the selection is done on the basis of evaluation of the quality or the significance of the earlier work. This will probably be followed by a discussion of the plan for the research to be reported: reasons will be given for thinking that the plan chosen was optimal, in the circumstances. That is an evaluative conclusion, based on reasoning; the very thing that was said to be a contradiction in terms. The results of the investigation are then reported, and its quality and the significance of the findings will be discussed; two more evaluative efforts. Evaluation is in fact the touchstone of scientific investigation. (Scriven, 1997a) (This and the following quotes referred as Scriven [1997a] or Scriven [1997b] do not appear in any published paginated format, but only exist in electronic, non-paginated versions; however, these have all been authenticated as accurate by Michael Scriven.)

What an appealing view of evaluation! Evaluation is everywhere in science. But it doesn't stop there. Scriven continues:

Turn to the rest of the scientific life of scientists, the very scientists who still deny the legitimacy of evaluation as a scientific procedure. They teach; and every student they teach, they evaluate. If those evaluations are challenged, they defend them with data and reasoning. That is, they think these evaluations are objective, which is quite a different matter from thinking them to be highly precise. They review work submitted to journals; doing this is pure evaluation. They assess applications for research funding; this is evaluation. They consider the applicants for a job in their department or research group; this is entirely an evaluative process. Not only does the intrinsic scientific process consist largely and crucially in evaluation, but the social operation of every scientific discipline is held together by a net of evaluations. In passing we might note that every scientist not only does and defends evaluation in his or her private life, but takes counsel from those with expertise in doing it, for example the consumer magazines and the reports on medicines done by federal agencies. (Scriven, 1997a).

Now we are getting close to the core of the problem. For Scriven, evaluation is just about everywhere. It pervades all aspects of science. It influences how we choose our car or our toothpaste. To an evaluator with concerns about the potential market for our trade, this is a consoling perspective. Our market is nearly boundless.

The problem with this view is that Scriven confuses the everyday prosaic use of the term evaluation—the Webster definition, if you will—with the discipline or profession of evaluation, to the extent that either really exists in definable form. What he is doing is analogous to saying that every child is in fact an experimental scientist because, in learning that a light switch turns on a light or that throwing your food on the floor elicits a specific parental response, they are using essentially the same methodology that experimentalists follow in the lab. It's an attractive view to those who promulgate scientific metaphors but, unfortunately, it's not an especially good way to define science. And Scriven's approach is not an especially good way to characterize evaluation. It is true that evaluative, critical thinking permeates everyday life. In this sense, we should be teaching good evaluative skills in elementary school. And we do, to some extent. My daughter's school has as one of its goals the idea of teaching critical thinking and reasoning skills. But I wouldn't say they are training "evaluators" except in the most prosaic sense. The insight that evaluative skills permeate everyday life does not serve us well when we are trying to define evaluation as a professional endeavor. We shouldn't fault Scriven for his excess in defining evaluation so abstractly that it encompasses nearly everything—after all, he has the baggage of a philosopher to carry. But we will need to push beyond philosophical abstraction if we are to develop a definition that makes sense for the evaluation profession.

Scriven's view does, however, force us to think about how evaluation, as the endeavor we know, is different from the everyday meaning of the term. I'm not going to try to define evaluation here—that seems more appropriately relegated to one of those extended EvalTalk threads. But I will point out a few ways I think we would all agree evaluation as a discipline may differ from the everyday use of the word. Evaluation, like the scientific enterprise generally, is an *interpersonal and organizational* endeavor. It involves many people—it typically can't be done by an individual operating in a vacuum. There are numerous stakeholders involved. Evaluation is a *political* endeavor. There are competing interests that have varying power to influence what we are observing and can elect to ignore us altogether. Evaluation is a *public* endeavor. It takes place in a public arena and, like applied social science or politics, what we do will be held up to public scrutiny and criticism. Evaluation is usually a *decision-*

*driven* endeavor. We seldom have the luxury of conducting research at the pace our social science cousins might prefer. Decisions will be made on the basis of the best available evidence and it is our job to provide that, typically under significant time and resource constraints. Evaluation is a *formal methodological* endeavor. When I evaluate which car I want to buy, I do not necessarily define the method I use to make my decision. I may, if challenged, defend my decision based on how I arrived at it, but I seldom propose a formal approach in advance for how I will go about making my decision. Whether quantitative or qualitative, constructivist or post-positivist, we evaluators take great pains in describing the *how* of what we do. Finally, evaluation is increasingly a *professional* endeavor. We have associations, annual meetings, career paths, and increasingly talk about things like accreditation and certification.

So, maybe I will attempt a definition, not one that will stand up to scrutiny, but one designed to include some of these crucial distinctive characteristics that are so often omitted from our definitions. Here goes:

Evaluation is a *profession* that uses *formal methodologies* to provide *useful empirical evidence* about *public entities* (such as programs, products, performance) in *decision-making contexts* that are inherently *political* and involve multiple often-conflicting *stakeholders*, where *resources* are seldom sufficient, and where *time-pressure*s are salient.

This will, of course, not satisfy. But it does a couple of things that other definitions do not. It describes something closer to what we actually do. It is inclusive of all of the different methodological and epistemological schools and camps. It distinguishes what we do from some of our closer cousins, in particular applied social scientists, accountants and auditors (notice that my definition doesn't mention anything about money). So, it meets Scriven's criterion of "demarcation" in setting what we do apart from other endeavors. And, it leaves out the phrase "worth or merit." Note, however, that it is not an entirely value-free definition—I did say we provide *useful* empirical evidence, and that inevitably requires us to define the value of usefulness.

As I said earlier, this is probably better as grist for the EvalTalk mill than as a durable definition of evaluation. But, it does illustrate my rejection of Scriven's everyday definition. His view, while appealing to evaluators, does not address well the real-world description of what we do and who we are.

### **Checklist Methodology**

The third point Scriven makes is that we need to revisit the value of checklists as methodological tools in our evaluative work. He distinguishes between item checklists—essentially descriptive lists—and criterion checklists. I think he makes some very important, useful points here. In advocating criterion checklists he reminds us of the multivariate nature of our work, and of how easy it is to overlook essential criteria when we are in the midst of complex evaluation contexts. He says:

In my experience, a large proportion of program evaluations suffer from missing one or another of the subject-specific considerations that they should have included, such as the failure to look at the long-term support problem when evaluating programs to computerize schools or colleges. (Scriven, 1997b)

We're like the airline pilots who need a checklist not so much because they don't know what they have to do, but because they may forget to do everything important when operating something as complex as a plane.

At first this idea struck me as simplistic. It's obvious that a checklist can help me to remember what needs to be covered. But, upon reflection, I think this simple idea may be more profound than it first appears. Especially when we think about specific substantive areas—the classroom, the factory floor, the drug rehabilitation agency, the hospital—the need for checklists of things we should address in an evaluation becomes more powerful. In some ways it serves to set us apart from our social science cousins. An economist might get away with looking at cost-benefit ratios. A psychologist might emphasize attitude and motivation. A sociologist might investigate social networks or communication patterns. But we evaluators will have to attend to all of these domains and more in doing our work. We need to develop multivariate criterion checklists of what to do and what to look for that are tailored to the substantive arena being studied and that are based on our cumulative experiences, including those mistaken prior evaluations that omitted important factors.

### My "Evaluation" of Scriven's Paper

It is wholly appropriate that I end my comments with my evaluative conclusions. Note that I refer here to the everyday Scriven-ish definition of evaluation as critical judgment, not to the kind of evaluative judgements we in the evaluation profession routinely make. After all, I haven't assessed the need for Scriven's presentation. I haven't conducted a survey of your reactions. I haven't measured the effect of his speech on subsequent EvalTalk postings or on the way we will practice evaluation in the future. In short, I haven't done what we might call an "evaluation" of his speech. But I will do what a layperson might understand as "evaluate" his speech, in the everyday sense of the term. In fact, I'll use a reporting method familiar to us all and popularized by the film critics Siskel and Ebert. (Note that I don't consider Siskel and Ebert evaluators, even though they make evaluative judgments).

To Scriven's notion of minimalist theory I give a strong thumbs-down. His plot is just too shrill and could benefit from a broader fleshing-out of the value that adding theoretical perspectives can bring to even causal assessment in evaluation. To his definition of evaluation I have to give a reluctant thumbs-down. While his everyday definition moved me emotionally, I just don't think it stands up to scrutiny in the light of day. Finally, I think his idea of checklist methodology is a very exciting one. We can all benefit from developing more detailed, substantive, experience-driven lists of what to include in our evaluation work. For that idea I give him two big thumbs-up!

### REFERENCES

- Chen, H. T. & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review*, 7, 283-302.
- Chen, H. T. & Rossi, P. H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning*, 10, 95-103.
- Scriven, M. (1997a). The new science of evaluation. Paper presented at the annual meeting of the Australasian Evaluation Society.
- Scriven, M. (1997b). Minimalist theory. The least theory that practice requires. Paper presented at the annual meeting of the American Evaluation Association.
- Scriven, M. (1997b). Minimalist theory. The least theory that practice requires. *American Journal of Evaluation*, 19(1), 57-70.
- Trochim, W. (1985). Pattern matching, validity and conceptualization in program evaluation. *Evaluation Review*, 9(5), 575-604.