

Evaluation Science

Michael Quinn Patton¹

American Journal of Evaluation
2018, Vol. 39(2) 183-200
© The Author(s) 2018
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1098214018763121
journals.sagepub.com/home/aje



Abstract

Culturally and politically science is under attack. The core consequence of perceiving and asserting evaluation as science is that it enhances our credibility and effectiveness in supporting the importance of science in our world and brings us together with other scientists to make common cause in supporting and advocating for science. Other consequences include communicating our role more clearly and credibly to those who value science. Viewing evaluation as science may affect how we are viewed, treated, and positioned in academia, government, and by funders and users of evaluation. The ramifications of evaluation science for evaluation's status as a profession, branch of applied social science, discipline, and transdiscipline are reviewed. The conclusion offers implications, caveats, and cautions regarding the identity and practice of evaluation science.

Keywords

science, discipline, transdiscipline, evaluation science, profession

I am an Evaluation Scientist. I do Evaluation Science.

On April 22, 2017, millions marched for science in 600 cities worldwide (Fleur, 2017). The American Evaluation Association (AEA) was one of the 270 partner organizations that supported the March for Science (<https://www.marchforscience.com/partners/>). I joined more than 40,000 in Washington, DC, marching with my son, daughter-in-law, and granddaughter. The questions of the day were, naturally enough: What brings you to the march? Are you a scientist? What kind? If not, what's your connection to science?

The first time I had this conversation was with a stranger after she successfully elicited a smile from my 9-month-old granddaughter. My daughter-in-law, a cell biologist at Yale, shared her identity without a moment's hesitation. I paused. I've been an evaluator for 45 years. I haven't identified myself as a sociologist in years. Saying I was a social scientist seemed vague and would naturally invite the follow-up query: What kind? So, I tried on a new identity. "I'm an evaluation scientist," I said. "I do evaluation science."

At first the phrase felt strange, awkward, even alien. I expect that I sounded tentative. And, of course, I had to explain what evaluation science is, which I got better at as the March progressed. But the real surprise was that, after several hours and multiple repetitions, I realized that the label

¹ Utilization-Focused Evaluation, St. Paul, MN, USA

Corresponding Author:

Michael Quinn Patton, Utilization-Focused Evaluation, 740 Mississippi River Blvd. South, Suite 21E, St. Paul, MN 55116, USA.
Email: mqpatton@prodigy.net

resonated. I liked proclaiming myself an evaluation scientist. I reveled in explaining evaluation science. I even formulated an “elevator speech” to explain evaluation science:

Science is systematic inquiry into how the world works. Evaluation science is systematic inquiry into how, and how well, interventions aimed at changing the world work.

Why This Matters

Both science in general and evaluation in particular are evidence-based processes with conclusions derived from systematic inquiry to understand and explain how some aspect of the world works. The credibility of scientific evidence is under attack. Guilt by association, the credibility of evaluation evidence is diminished. To defend the value of scientific evidence, then, is to defend the value of evaluation evidence. It is in our interest as evaluators to make common cause with those who support science.

The antiscience barbarians are not just at the gates, they have entered and taken control of the castle. That concern is what motivated the Marchers for Science. The *New York Times* headline on the day of the March read: *Scientists, Feeling Under Siege, March*. The *USA Today* led with: *Marchers for Science protest “Alarming” Antiscience Trends*.

Culturally and politically the antiscience trends include “alternative facts,” “fake news,” and a “posttruth” world. In November 2016, the Oxford Dictionaries announced posttruth as its international Word of the Year.

Posttruth

Relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief.

Casper Grathwohl, President of Oxford Dictionaries, explained: “Given that usage of the term hasn’t shown any signs of slowing down, I wouldn’t be surprised if posttruth becomes one of the defining words of our time.”

The implication of living in a posttruth political culture is that science becomes just another perspective. Scientific evidence is no more valid than personal opinion. Every political persuasion advocates not just values but promulgates its own “facts.” The distinction between evidence and opinion becomes blurred. This, by extension, has a corrosive and delegitimizing effect on evaluation. Evaluation findings become just another kind of opinion. Truth (Blackburn, 2005), finding and speaking it, is devalued and disputed.

Beyond assertions that climate change is a hoax and doubts about the safety of vaccines, antiscience trends include budget cuts for scientific research, including agencies like the National Institutes of Health and the Environmental Protection Agency. So, there were March for Science signs that shouted, “KEEP NIH HEALTHY,” “PROTECT the EPA,” and

E = Evidence

P = Protects

A = All of us.

I saw no signs that read “Protect Evaluation Units.” But I remember the Reagan Administration (1980–1988) systematically shutting down evaluation functions across the federal government. Funding for evaluation is ever vulnerable. The current antiscience political climate calls us to unite with others engaged in defending and supporting science, creating a united front to the larger world.

In what follows I'll consider the case for Evaluation Science. This will require examining what science is, what evaluation is, and thereby, what evaluation science is, or might be. I'll also address arguments to the contrary, that evaluation is not science (or at least some approaches to evaluation do not so qualify). As context, I'll consider how thinking of evaluation as science complements but is different from thinking of evaluation as a profession and discipline.¹

What Is Science?

[Science is] the intellectual and practical activity encompassing the systematic study of the structure and behavior of the physical and natural world through observation and experiment. (Science, n.d., def. 1).

[Science is] a systematically organized body of knowledge on a particular subject. (Science, n.d., def. 1.2)

Science: systematized knowledge derived from observation, study, and experimentation carried on in order to determine the nature of principles of what is being studied. (*Webster's New World Dictionary*, 1982, p. 1275)

The Random House unabridged dictionary defines science as

A branch of knowledge or study dealing with a body of facts or truths systematically arranged . . . ; Systematized knowledge in general; knowledge, as of facts or principles; knowledge gained by systematic study; a particular branch of knowledge. (Stein & Urdang, 1966)

Science historian David Wootton (2015) makes the case that

modern science was invented between 1572, when Tycho Brahe saw a nova, or new star, and 1704, when Newton published his *Optiks*, which demonstrated that white light is made up of light of all the colors of the rainbow, that you can split it into its component colors with a prism, and that color inheres in light, not in objects. (p. 1)

He identifies astronomy as the first modern science:

What made astronomy in the years after 1572 a science? It had a research programme, a community of experts, and it was prepared to question every long-established certainty . . . in light of new evidence. Where astronomy led, other new sciences followed. (Wootton, 2015, pp. 1, 2)

Light is a phenomenon of nature. It exists independent of human thought. Science, in contrast, is a human invention, a social construction. "Science is a way of thinking much more than it is a body of knowledge." Carl Sagan, cosmologist and science popularizer (<https://www.singularityweblog.com/carl-sagans-last-interview-science-as-a-candle-in-the-dark/>). Science, then, requires scientists.

A scientist is defined as an expert in science, according to the Random House unabridged dictionary (Stein & Urdang, 1966). A scientist, it turns out, is a person who works in and has expert knowledge of a particular field of science, or, more generally, any person who studies or works in a scientific field. The *Oxford English Dictionary* dates the origin of the word "scientist" to 1834. Still scientists themselves vary in how they describe science.

What Is Evaluation?

Evaluation involves making judgments about the merit, value, significance, credibility, and utility of whatever is being evaluated, for example, a program, a policy, a product, or the performance of a person or team. That is at least one perspective on what evaluation is. Defining evaluation turns out

to be challenging. Evaluation, like science, is a social construction. Evaluation, within a given context, is whatever a group of people agree that it is. The mission of the AEA is to “improve evaluation practices and methods, increase evaluation use, promote evaluation as a profession, and support the contribution of evaluation to the generation of theory and knowledge about effective human action” (2016). To promote evaluation, the AEA needs to be able to explain what evaluation is, as do member evaluators.

But there’s the rub. Evaluators are an eclectic group working in diverse arenas using a variety of methods drawn from a wide range of disciplines applied to a vast array of efforts aimed at improving the lives of people in places throughout the world. Evaluators don’t all define evaluation the same way. Indeed, the field of evaluation has a history of vociferous debates about how to define evaluation, what methods to use, how to judge quality, and what competences are needed to be an evaluator, to name but a few of the contentious issues discussed and debated. Still, despite different perspectives and rough encounters among those with opposing views, an association promoting evaluation needs to elucidate what it is.

To that end, I was asked in 2010 by the board of the AEA to chair and facilitate a task force of diverse evaluators to produce a statement explaining evaluation. Seven AEA members, including an AEA staff representative, worked on this challenge for 14 months, from March 2010 to May 2011. My responsibility as facilitator was to bring together diverse perspectives, find common ground, draft a document for review, incorporate feedback, and work with task force members to come up with a statement. After months working together, we succeeded. (For details on this process, see Patton, 2018, pp. 6–17.)

In submitting our statement to the AEA board, we recommended that the document be treated as a resource for AEA members and staff but that it not be officially adopted or endorsed by the board or the membership. Rather, we suggested it be posted on an AEA blog-like site, so members could add comments and exchange views on the document and otherwise evaluate the statement. The AEA board accepted our recommendation. We committed to periodically revising the *What Is Evaluation* statement based on comments and developments in the field. The first revision since the original version was presented at the 2017 annual conference of the AEA in keeping with the theme: *From Learning to Action*. The main revision was adding a paragraph on alternative perspectives on evaluation that includes Evaluation Science. This article explains and justifies that addition, which is reproduced at the end of the next section on Defining Evaluation Science. The full *What Is Evaluation* statement remains open for dialogue, comment, critique, and applause on the AEA website (<http://www.eval.org/p/bl/et/blogid=2&blogaid=4>).

Evaluation Science

Earlier, I presented dictionary definitions of science. There are variations on the precise wording in different dictionaries, but the consistent theme is systematic inquiry to understand and explain how some aspect of the world works. Evaluators do that. Indeed, the first AEA guiding principle on *Systematic Inquiry* states “Evaluators conduct systematic, data-based inquiries about what is being evaluated.”

The Science Council defines Science as “the pursuit and application of knowledge and understanding of the natural and social world following a systematic methodology based on evidence” (<http://sciencecouncil.org/about-us/our-definition-of-science/>). Data and evidence are foundations of evaluation practice. Common signs at the March for Science were “Evidence NOT ideology” and “Data NOT dogma.”

Defining Evaluation Science

Based on common understandings, perceptions, wisdom about, and definitions of science and evaluation, let me proffer a working definition of evaluation science that incorporates the explanation I had the opportunity to field test during the March for Science.

Science is systematic inquiry into how the world works. Evaluation science is systematic inquiry into how, and how well, interventions aimed at changing the world work. Evaluation science involves systematic inquiry into the merit, worth, utility, and significance of whatever is being evaluated by adhering to scientific norms that include employing logic, using transparent methods, subjecting findings to review, and providing evidence and explicit rationales to support reason-based interpretation, valuing, and judgment. (This is the paragraph added to the AEA website statement as a comment on *What Is Evaluation*.)

Evaluation Science as a Disciplinary Body of Knowledge

Different scientific disciplines are distinguished by the overarching questions they ask and the body of knowledge that is built through inquiry into those questions. Evaluation as a scientific specialization asks: *What are the factors that contribute to, methods for determining, and criteria for judging interventions' successes and failures?* Interventions are any effort, program, project, initiative, product, policy, organization or community development, or activity aimed at bringing about change. An expert evaluator knows how to conduct an evaluation of a particular intervention in a particular context for a particular purpose. As an evaluation scholar in *the discipline of evaluation*, a knowledgeable evaluator contributes to and has access to *a body of knowledge* about ways of studying and judging interventions and ways of applying knowledge to design and improve interventions, both based on empirically and theoretically validated patterns of successes across interventions and evaluations. It is because *evaluation has become a reservoir of knowledge about effectiveness* that we are consulted about how to design, plan, and implement new interventions, not just evaluate them once implemented.

In their classic *Foundations of Program Evaluation: Theories of Practice*, Shadish, Cook, and Leviton (1991) examined evaluation as a methodological specialty and a profession of practice, but it was evaluation theory made coherent by a validated body of knowledge that made evaluation a scientific discipline.

Program evaluators are slowly developing a unique body of knowledge that differentiates evaluation from other specialties while corroborating its standing among them. Evaluation is diverse in many ways, but its potential for intellectual unity is . . . “the logic of evaluation,” which might bridge disciplinary boundaries separating evaluators. (p. 31)

They identified five fundamental issues that are at the core of evaluation's disciplinary body of knowledge:

1. *social programming*: ways that social programs and policies develop, improve, and change, especially in regard to social problems
2. *knowledge construction*: the ways researchers learn about social action
3. *valuing*: the ways value can be attached to program descriptions
4. *knowledge use*: the way social science information is used to modify programs and policies
5. *evaluation practice*: the tactics and strategies evaluators following the professional work, given the constraints a face. (p. 32)

The body of knowledge accumulated around these five conceptual issues constitutes evaluation's distinctive theory. Specific approaches to addressing and understanding these issues are what

distinguish particular evaluation theories. More than a decade later, Alkin and Christie (2004) created an evaluation theory tree providing both the conceptual framework and image for evaluation's body of knowledge based on attention to evaluation use, evaluation methods, and valuing—the three branches of the theory tree. They subsequently expanded the theory tree with recent and international contributors to evaluation's body of knowledge (Alkin, 2013).

Stufflebeam, Madaus, and Kellaghan (2000), Madaus, Scriven, and Stufflebeam (2012), and Stufflebeam and Coryn (2014) are among many other examples of efforts to document, capture, synthesize, and report evaluation's rapidly evolving body of knowledge represented by diverse viewpoints, models, methods, and approaches. The scientific discipline of evaluation has handbooks (Guttentag & Struening, 1975; Shaw, Greene, & Mark, 2006; Newcomer, Hatry, & Wholey, 2015), an encyclopedia (Mathison, 2005), peer-reviewed journals (e.g., *American Journal of Evaluation*, *New Directions for Evaluation*, *Evaluation*, *Evaluation Review*, and *Evaluation and Program Planning*). Our scientific disciplinary stature is solid and growing.

Pioneering evaluation thought leader and philosopher Michael Scriven posits four criteria to merit scientific disciplinary status:

1. a distinguishable subject matter,
2. subject-specific methods,
3. substantial field of application, and
4. production of results serves substantial social and/or intellectual improvement.

He finds that evaluation meets all these criteria (Scriven, 2004, p. 186). In his masterful and enduring *Hard-Won Lessons* treatise (Scriven, 1993), he sings such praises of evaluation as a scientific discipline as to make a shy evaluator blush, to wit:

Evaluation, the ugly frog, turns into a prince when the spell on us is broken. Evaluation is all that distinguishes astronomy from astrology, good explanations from bad ones, good experimental designs—or bridge designs—from inferior ones, good scientists and engineers and technologists from run-of-the-mill ones. It is a discipline that is part of every discipline because it distinguishes the discipline from pretentious jargon, just as one distinguishes good food from garbage. (pp. 47, 48)

Scriven pioneered treating evaluation as the Science of Valuing (Shadish, Cook, & Leviton, 1991, p. 74). He has reflected that “much of my early work in evaluation was devoted to attacking various attempts to . . . marginalize evaluation into the category of decision-support apparatus, *by contrast with* truth-seeking efforts of “true science” (Scriven, 2004, p. 188, emphasis in the original). For Scriven, evaluation is without question “*scientifically legitimate* . . . The bottom line . . . is that a competent program evaluator can show that, for example, a program for teaching reading is truly excellent or truly worthless *as a matter of scientific fact*” (Scriven, 2013, p. 171, emphasis in the original).

Here, as with evaluation's professional status, academics will argue among themselves about whether, and the extent to which, evaluation is a scientific discipline. Arguing about definitions and boundaries is what academics do, and I've had my share of fun so indulging. But, to the outside world, when I am talking about evaluation, I proclaim, assert, and celebrate without reservation evaluation's disciplinary status as a theory-based body of scientific knowledge. Our peer-reviewed journals capture, adjudicate, and report that body of knowledge. The nature and depth of our disciplinary status may be a matter for academic debate but the fact that we have a reservoir of scientifically based theory and knowledge should, from my perspective, be part of our public persona and identity to the larger world.

Evaluation Science as a Transdiscipline

More recently, Scriven (2008) eloquently and forcefully envisions and advocates positioning evaluation as the *alpha transdiscipline*: As a transdiscipline, evaluation sits atop the disciplinary, scholarly, academic, and scientific hierarchy with philosophy, logic, and statistics as bodies of knowledge, theory, and methods that are essential for the scholarship, knowledge creation, and scientific rigor of all other disciplines. Evaluation is the *alpha* scientific transdiscipline because “its domain includes the methodology of the task of validation at any discipline’s claim to legitimacy as a discipline: it is the master of credentials” (Scriven, 2013, p. 175).

Scriven does not aspire to have evaluation become more like science, for he has long been scathing in his critique of shoddy reasoning and judgment in much scientific practice (Scriven, 1976). Rather, he challenges science to become more evaluative. Indeed, he is critical of evaluators for failing to recognize evaluation’s preeminent scientific status.

Some leading texts on evaluation describe evaluation as a branch of social science, which is surely not true since the social sciences have no methodology for dealing with evaluative propositions, and are only partially recovered from the days when they put all evaluation in the untouchable class. I would like to see the exact reverse of this relationship come to pass [Instead of] the view of evaluation fields as being satellites circling social science’s sun. I think a more appropriate model is the reverse. (p. 20)

As for scientists’ failure to recognize evaluation as science, his eloquence resists summary:

The status of evaluation in the twentieth century represents one of the most striking paradoxes in the history of thought: An essential—and perhaps the most important—ingredient in all intellectual and practical activity has been explicitly banned or implicitly excluded from discussion or acknowledgment in most of its natural territory.

The psychological, social, and political reasons, as well as the intellectual ones, for this bizarre situation—for this “intellectual treason of the intellectuals” . . . [has come with enormous cost] If the status of evaluation as an “untouchable” subject is to be radically changed, there must be a direct attack on the myths responsible for this paradox, and there must also be works that set out the elements of an alternative view of evaluation, a view of it as perhaps the most important and pervasive process in which the human mind is capable. (Scriven, 1991, p. 10)

You can see how Scriven builds on that judgment to assert evaluation’s rightful status as the α transdiscipline of all sciences, which means it must, logically, be a science.

. . . . [W]hat science should be is an evaluative decision, and it is hard to argue that there is a more important kind of decision anywhere in science. This is the decision, or a refinement of it, after all, that we call on many of our peer review panels to make. In particular we ask them to review the scientific *quality* of the proposals in front of them. They are indeed evaluators, and on them and on their peers on editorial boards and appointment committees rest the whole definition and quality of science, and eventually the whole future of science—and of mathematics, and engineering, and technology. . . . [S]cience is at its core as well as in all of its applications an evaluative enterprise.

If science is an evaluative enterprise, then where in the whole of science and scientific training is the training in evaluation? It is not there. (Scriven, 1993, p. 48)

Other Foundations of Evaluation as Science

Stewart Donaldson’s (2007) book on *Program Theory-Driven Evaluation Science* defined the term as follows:

Evaluation science (instead of evaluation) is intended to underscore the use of rigorous scientific methods (i.e., qualitative, quantitative, and mixed-method designs) to attempt to answer valued evaluation questions. A renewed emphasis in evaluation practice on relying on systematic, scientific methods is especially important for overcoming the profession's negative reputation in some contexts. That is, in some settings, evaluation is criticized for being an unreliable, soft, or second-class type of investigation. The term *evaluation science* signals the emphasis placed on the guiding principle of systematic inquiry (Guiding principles for evaluators, 2004) and the critical evaluation standard of accuracy (joint committee on standards for educational evaluation, 1994; p. 11, emphasis in the original)

Donaldson's book provides importance guidance for the practice of evaluation science, especially highlighting program theory as the centerpiece of such practice. He began deliberately using the terminology "evaluation science" as he encountered evaluation in academic settings being treated as second-class, not really science. Since having taken to referring to what he does as evaluation science, he reports gaining credibility and greater acceptance among social and behavioral scientists with whom he interacts (Donaldson, 2017). The issue here is our identity as much as our practice.

The Science of Evaluation: A Realist Manifesto by Ray Pawson (2015) offers another window into the practice of evaluation science. He examines whether evaluation has produced a cumulative and authoritative body of knowledge deserving of the designation evaluation science. He provides a detailed blueprint for an evaluation science based on realist principles. Although his primary concern is with making the case for realist evaluation as the primary legitimate and credible approach to evaluation science, his positive experience explaining evaluation inquiry as worthy of scientific designation merits our attention.

The Ascendance of Scientific Designation as a Source of Legitimacy

Language matters. Evaluation language matters.

Rodney Hopson (2000)

Precedent-setting examples of branding a field as science include policy science (Lasswell, 1970) and action science (Argyris, Putnam, & Smith, 1985). New fields of inquiry and practice are flourishing—and asserting their legitimacy by taking on the moniker of science:

- complexity science (<http://www.complexity.ecs.soton.ac.uk/>),
- cognitive science (<http://www.cognitivesciencesociety.org/>),
- improvement science (Christie, Inkelas, & Lemire, 2017),
- implementation science (Eccles & Mittman, 2006),
- translational science (Pardridge, 2003),
- urban science (Townsend, 2015),
- management and service science (MASS, 2017),
- the sciences of learning and instructional design (Lin & Spector, 2017),
- developmental science (Nelson, de Haan, & Quinn, 2016),
- community science (Chavis, 2015), and
- big data science (<https://datascience.berkeley.edu/what-is-big-data/>).

Viewing evaluation science as the alpha transdiscipline à la Scriven, these could all be considered specializations within evaluation science, and *Evaluation Science* might well be the umbrella under which they all operate.

A Comparative Example in Greater Depth

Let's take a closer look at one such new specialty, *sustainability science*, for similarities to evaluation science and one notable difference. Sustainability science seeks to facilitate what the National Research Council has called a "transition toward sustainability," improving society's capacity to use the earth in ways that simultaneously "meet the needs of a much larger but stabilizing human population, . . . sustain the life support systems of the planet, and . . . substantially reduce hunger and poverty." Sustainability science, like evaluation, has developed from an emergent research field into a vibrant discipline with scientific conferences, journals and an association. Like evaluation, the field is characterized more by its purpose, elucidating and supporting sustainability, than by a common set of methods (Spangenberg, 2011, p. 175). There is one way, however, in which sustainability science is different from evaluation science. Sustainability science has been recognized as a scientific specialization by the National Academy of Sciences.

In early 2005, Bruce Alberts and Ralph Cicerone, in their respective roles as outgoing and incoming presidents of the National Academy of Sciences, proposed that the maturing field of sustainability science might be ready for a "room of its own" in PNAS. After a committee study and extended discussion, the PNAS Editorial Board approved a new section on Sustainability Science, which now shares the masthead with other long-term residents such as Physics, Genetics, and Cell Biology. (Clark, 2007, p. 1738)

I believe that AEA should seek designation of evaluation as a science with the National Academy of Sciences in the appropriate applied social sciences division.

The thrust of this review has been to bring *evaluation as science* more fully into the consciousness and identity of evaluators, the scientific community, and the larger world of evaluation stakeholders. I conclude with some implications, caveats, and cautions about taking on the mantle of evaluation scientist. These are issues that have arisen in my personal inquiry into evaluation science since participating in the March for Science.

Implications, Concerns, and Caveats

Scientism and the "Scientific Method"

Scientific inquiry is not limited to a narrow definition of the scientific method. Evaluation science is, likewise, not reducible to, defined by, or limited to certain preferred methods. As I began discussing the idea of evaluation science with colleagues and offered an AEA eStudy course on evaluation science (June 2017), the first reaction I received was that I must be talking about ensuring that evaluations used "rigorous" methods to qualify as "scientific." I am not going down that road, though some reviewers of an earlier version of this article urged me to do so. My position is that science involves systematic inquiry and rigorous thinking, but that rigor does not reside in methods per se. Making that case, which is important for conceptualizing evaluation science, requires a bit of history.

The *Wikipedia* entry on science states that during the 19th century, the word "science" became increasingly associated with the scientific method, which led to the establishment of scientific institutions where the method could be supported. In their classic review of *Foundations of Program Evaluation*, Shadish et al. (1991) observed that "as a specialty, evaluation is most like methodological specialties—ethnography, psychometrics, experimental design, or survey research" (p. 31).

The most widely used textbooks and sourcebooks in program evaluation deal primarily with methods . . . The popularity of methodological books is no surprise, given the program evaluation is a pragmatic activity. Evaluation practitioners need to act and need tools to use in their daily work. Methods texts are their essential references. (p. 34)

Rossi, Freeman, and Wright (1979), in the leading methods textbook, *Evaluation: A Systematic Approach*, defined evaluation research as the systematic application of social research procedures in assessing social intervention programs. The book, now in its seventh edition (Rossi, Lipsey, & Freeman, 2003), positions evaluation as applying social science methods. Rigorous methods, in this positioning, define evaluation quality. One of the central issues that emerges in focusing on evaluation methods as the defining characteristic of the field is whether those methods are sufficiently rigorous to be considered worthy of the designation science. Science, in this tradition, is often equated with the scientific method.

When conducting research, scientists use the scientific method to collect measurable, empirical evidence in an experiment related to a hypothesis (often in the form of an if/then statement), the results aiming to support or contradict a theory. (Bradford, 2015, p. 1)

The Scientistic Ideal of Evaluation circa 1980—and now

Scientistic, characterized by or having an exaggerated belief in the principles and methods of science. (*Random House Dictionary*, Stein & Urdang, 1966)

The Stanford Evaluation Consortium was established at Stanford University in 1974 under the leadership of distinguished evaluation pioneer Lee J. Cronbach. Each year through 1979, some 20 faculty members and an equal number of doctoral students discussed major evaluation issues. The results of their ruminations were published in a book entitled *Toward Reform of Program Evaluation* (Cronbach and Associates, 1980), which featured 95 theses (evoking Martin Luther). It remains a provocative and stellar list. Let me begin at the end with the final thesis.

#95. Scientific quality is not the principal standard; an evaluation should aim to be comprehensible, correct and complete, and credible to partisans on all sides. (p. 11)

This concluding thesis was a reaction against the predominant view at the time that “the job of evaluation is said to be scientific appraisal, and the conventional presentation tells the evaluator and the purchaser of evaluative services to judge evaluations by their rigor” (p. 54). Rigor meant randomized controlled trials. That was the position, for example, of the leading textbook, *Evaluation: A Systematic Approach* (Rossi, Freeman, & Wright, 1979), as well as many other sources Cronbach cites in a 10-page critique of *Evaluation Defined as Scientific Appraisal* (pp. 55–64). Toward the end of the book he returns to this critique.

Unfortunately, criticism [of evaluations] in recent years have been predominately scientistic. Only a numerical index of a “treatment effect” that has been tested for “significance” is seen as a worthy and product. This view of what an evaluation offers is distressingly limited, and with respect to some programs and contexts is wholly inappropriate. Scientistic commandments for evaluation oversimplify. (Cronbach et al., 1980, p. 332)

Cronbach (1982) takes up the limitations of “the scientistic ideal” in his subsequent groundbreaking book on *Designing Evaluations of Educational and Social Programs* in which he takes on issues of causality, replication, standardization, and generalization and redefines what rigor means for evaluation consistent with thesis 95 cited earlier.

But judging whether evaluation can be viewed as science based on rigor, where rigor is defined by use of experimental methods and research designs that scientistically address causality, replication, and generalizability, remains widespread. In discussions about positioning evaluation as science, the primary argument of naysayers is that evaluation, in general practice, lacks sufficient rigor

to be considered science. Yet a considerable literature, building on Cronbach's seminal works, demonstrates that rigorous science involves much more than experimental methods and that evaluation can and ought to employ a variety of methods appropriate to the situation. In a thoughtful and insightful *Encyclopedia of Evaluation* (Mathison, 2005) entry on methodological debates in evaluation, distinguished evaluation practitioners and methodologists Jennifer Greene and Gary Henry (2005) concluded:

If the standards for what constitutes legitimate evidence relevant to policy decisions and program continuation narrow too much, excluding evidence from all but large-scale, randomized experiments, we will obtain only very limited information on very few programs A case in point is the federal government's contemporary push for "scientifically based evidence," favoring especially experimental evidence The contextual insights, program stories, and participant feedback available from case studies, in-depth interviews, surveys and other methods will be omitted. Moreover, narrowing the types of evidence that are considered legitimate for program improvement and policy actions will silence the voices of the many program stakeholders who can and should be heard. We, quantitative and qualitative evaluators alike, should unite in worry that the absence of evidence that meets some narrowly drawn standard will become a license for actions based entirely on ideology or the force of unconstrained rhetoric (p. 350; for enhanced readability, the first two sentences have been reversed in this excerpt)

Narrow scientific criteria as a license for actions based entirely on ideology or the force of unconstrained rhetoric? They could hardly have been more prescient. The antiscience argument is essentially scientific: There is no definitive, completely agreed-on, absolute scientific proof that smoking causes cancer, or that vaccinations don't cause autism, or that humans cause climate change, or that racism is associated with police killings, or . . . , or . . . , or

As an antidote to narrow scientism, Greene and Henry (2005) call for evaluators to . . .

unite in our commitment to enact our hard-won acceptance of multiple methods and multiple ways of knowing, to reclaim the conversation about the contributions of social science to social policies and programs and refocus it on substance in values rather than on method, and thereby to redirect our collective evaluation expertise and energy in the service of democratic social betterment and social justice. (p. 350)

Thus, let me absolutely clear that asserting evaluation's scientific status is not aimed at advocating a narrow view of methodological rigor as residing in certain favored designs. As the writings on evaluation science of both Donaldson (2007) and Pawson (2013), not to mention Patton (2008, 2012, 2015), make clear, evaluation science is methodologically eclectic, pluralistic, and mixed. I reiterate that rigor resides in rigorous thinking not methods. In that regard, consider this assertion by Percy Bridgman, recipient of the 1946 Nobel Prize in Physics

There is no scientific method as such, but the vital feature of the scientist's procedures has been merely to do his utmost with his mind, no holds barred.

Evaluation Practice That Is Not Scientific

Some evaluation-related activities, like routine monitoring, internal improvement-oriented feedback for learning, accountability checklists and reporting, and unpublished proprietary evaluations may not meet the criteria for being engaged in scientific inquiry. These practices are, however, applications of evaluation science. They are based on and apply evaluation's knowledge base and methods. A parallel perspective emphasizes the technological nature of medicine. In her highly respected *History of Medicine*, Jacalyn Duffin (2010) asserts: "Medicine is not a science; rather, it is an applied

technology or an art that makes extensive use of science” (p. 65). Morell, distinguished recipient of the American Evaluation Association’s Lazarsfeld Award for Contributions to Theory, argues that “evaluation is far more of a technological than a scientific pursuit” (1979, p. 1). Jane Fields and Tim Sheldon (2017) used a technological lens to communicate about their evaluation of EngrTEAMS, a 5-year, US\$8 million project funded by the National Science Foundation, designed to increase students’ learning of science content, as well as mathematical concepts related to data analysis and measurement.

With EngrTEAMS, we recognized early on that the steps in an evaluation cycle are closely aligned to the steps of the engineering design process Focusing the evaluation design is similar to the planning stage in the engineering design process. Gathering credible evidence, justifying conclusions, and using the lessons learned are similar to the solution phases in the EngrTEAMS design process of trying, testing, and deciding. When we were able to demonstrate that the problem-solving process used in evaluation was similar to that used in their engineering design framework, program staff easily saw the connections and were even more receptive to the evaluation side of the work. (p. 1)

Morell’s full rationale for viewing evaluation as technology is available online (Morell, 1979) and provides a comprehensive review of the implications of a technological versus scientific identity and practice for evaluators. Interestingly, science and technology are so closely associated in the public mind, and apparently among policy makers, that the American Association for the Advancement of Science (AAAS) integrates science and technology in its mission statement:

The AAAS seeks to “advance science, engineering, and innovation throughout the world for the benefit of all people.” To fulfill this mission, the AAAS Board has set the following broad goals (partial list)

- Enhance communication among scientists, engineers, and the public;
- Strengthen support for the science and technology enterprise;
- Strengthen and diversify the science and technology workforce;
- Foster education in science and technology for everyone; and
- Increase public engagement with science and technology (AAAS, 2017)

Generating evaluation knowledge and developing theories and methods, then, can be thought of as evaluation science while *applying* evaluation knowledge, theories, and methods can be considered evaluation technology. Both designations position evaluation as more than an administrative, management, compliance, and accountability function. Evaluations that are mechanistically designed and mindlessly implemented to meet a compliance mandate are neither science nor technology. They are what Peter Drucker (1959) dubbed “knowledge workers” who produce reports instead of manufactured goods.

Avoiding Scientific Elitism

The demonstrably true statements of the sciences which, especially in recent times, had the uncomfortable inclination never to stay put, although, at any given moment they are, and must be, valid for all.

Philosopher Hannah Arendt (quoted in Peter, 1977, p. 459)

Being perceived as “valid for all” can be a challenge when scientists are perceived as distant and inaccessible, hiding out in the ivy tower or research lab. Some view scientists as elitist, arrogant, and aloof. Of course, some view evaluators the same way. Thus, being an evaluation scientist can come across as a double dose of elitism. We have worked to make evaluation understandable, practical, accessible, and useful. We have also been working to make evaluation diverse and inclusive. What

we have learned from those efforts can inform how we communicate about and practice evaluation science, including “seeking the unheard voices in science” (Olmstead, 2017). Stewart Donaldson explained in the evaluation science webinar that while he finds positioning evaluation as science enhances credibility in academia, he largely avoids that labeling when working with nonacademics in schools and community settings.

Citizen Evaluation Science

Science must enter into the consciousness of the people.

Albert Einstein, opening day World’s Fair, April 30, 1939 (quoted in Miner & Rawson, 2000, p. 429)

One way to avoid scientific elitism is to promote citizen evaluation science. I recently asked a group I was facilitating whether they’d prefer to be known as evaluation stakeholders or citizen evaluation scientists. The question provoked an animated discussion and concluded with an openness to try on the new moniker. Collaborative, participatory, and empowering approaches to evaluation can become collaborative, participatory, and empowering approaches to evaluation science. Check out the citizen science movement supported by *Scientific American* (<https://www.scientificamerican.com/citizen-science/>).

Evaluation Science Infused with Moral and Ethical Discourse

Science without conscience is the death of the soul.

Montaigne, *Essays 1580-88* (quoted in Miner & Rawson, 2000, p. 430)

There have been some concerns that science, and guilt by association, evaluation science, excludes or marginalizes social justice concerns. The March for Science was criticized as racist and sexist.

For the past three months, the scientific community, which is largely white, heterosexual, cisgender, able-bodied and male, has been fiercely debating the political nature of the march in the face of a Trump regime, leaving scientists from marginalized backgrounds feeling . . . well, further marginalized. In response, scientists who identify as women, disabled, queer, trans, people of color, etc. converged around the hashtag #MarginSci to take their racist and sexist colleagues to task. (<http://www.theroot.com/margin-sci-the-march-for-science-as-a-microcosm-of-lib-1794463442>)

This critique makes it critical to consider whether making social justice a scientific priority and achieving social justice a focus of scientific evaluation inquiry might well enhance attention to and understanding of the role of knowledge and science in furthering the cause of social justice and combatting racism and sexism. For a scientific argument in support of social justice, see Denzin’s (2014) *Qualitative Manifesto*. Although focused on qualitative inquiry, his fundamental argument for science in support of societal change and social justice is generally applicable.

David Chavis has designated his evaluation approach as “community science” (<http://www.communityscience.com/>) with an explicit social justice agenda. He specifically chose to associate his social justice commitment with science to communicate his commitment to conducting evaluations rigorously. Since he uses participatory evaluation approaches, I asked David if he had explored citizen science evaluation. He had not, but, in the context of the March for Science, he commented that “It’s about time that scientists saw themselves as citizens.”

Evaluation science then can and should incorporate moral discourse. Reference to evaluation science is not meant to reinforce a narrow technoscientific approach. Evaluation science must be concerned about both doing things right and doing the right things. Tom Schwandt’s (2002) call for moral discourse in evaluation and Scriven’s call for an ethical infusion as evaluation’s third

revolution take on ever greater relevance in a world where ideological antisience forces threaten to undermine both morality and ethics.

Evaluation science, truth, and quality. The oft-asserted evaluation mandate to speak truth to power takes on a renewed sense of urgency in a posttruth era and broadens to include speaking evaluation science truth to the general public. The credibility of science to seek and speak truth depends on quality control. Neither science, nor scientists, are inherently good. “Bad science” (Goldacre, 2009) is all too common and must be exposed and corrected to maintain scientific integrity. *Pandora’s Lab* (Offit, 2017) tells seven stories of “science gone wrong,” when purported scientific truths are wrong. *Rigor mortis* (Harris, 2017) examines “how sloppy science creates worthless cures, crushes hope, and wastes billions.” And how does one recognize, detect, and expose bad and sloppy science? Evaluation of science! And how does one recognize, detect, and expose bad and sloppy evaluation science? Meta-evaluation, the fifth and newest of the Joint Committee Standards.

Support for Evaluation Science Dialogue

You cannot hope to build a better world without improving the individuals. To that end each of us must work for his own improvement, at the same time share a general responsibility for all humanity, our particular duty being to aid those to whom we think we can be most useful.

Marie Curie (quoted in Partnow, 1977, p. 306)

We know from our own history of paradigm wars that hardening of the categories and talking past each other is not helpful. Not everyone who has doubts about climate change being human-caused is antisience. Not all parents who fear that a vaccination may have devastating side effects are antisience (<http://ngm.nationalgeographic.com/2015/03/science-doubters/achenbach-text>).

It is important not to demonize, ridicule, or alienate people who have doubts about certain specific claims of scientific knowledge. Science has been used for ill as well as good. What science proclaimed as true at one moment has been subject to revision. Carol Weiss (2002) called on evaluators to display appropriate humility in our claims and modesty in our interpretations. An invitation to take on the identity of evaluation scientist is not an invitation to deepen divisions but rather to expand dialogue.

Advocate for Science, Including Evaluation Science

The most remarkable discovery made by scientists is science itself.

Jacob Bronowski, *A Sense of the Future* (1977)

While you’re deliberating on the implications and connotations of *evaluation science*, visit the website for the AAAS, the world’s largest scientific organization. The AAAS seeks to “advance science, engineering, and innovation throughout the world for the benefit of all people.”

The principles and goals of the March for Science are congruent with a vision of evaluation science built on AEA’s mission and Guiding Principles (<https://www.marchforscience.com/mission-and-vision/>).

- Science that serves the common good
- Evidence-based policy and regulations in the public interest
- Support cutting-edge science education
- Enhance diversity and inclusion in STEM
- Engage in open, honest science, and inclusive public outreach

- Support funding for scientific research and its applications
- Humanize science
- Partner with the public
- Advocate for open, inclusive, and accessible science
- Support scientists
- Affirm science as a democratic value

To advocate effectively for science, including evaluation science, you may want to update your knowledge of the history and philosophy of science. If your last (or only) knowledge of science is Kuhn's classic, *Structure of Scientific Revolutions* (1970), it may be time for an update. A good place to start is Wootton's (2015) *The Invention of Science: A New History of the Scientific Revolution*.

From Evaluators to Evaluation Scientists

In closing, let me place the issue of who we are, what we do, and how we identify ourselves within a larger historical context. When I was president of AEA in 1988, the application to join the association asked for primary area of specialization: education, health, psychology, sociology, economics, criminal justice, so forth . . . Evaluation was not among the options listed. Our primary identity at the time was in our degree field with evaluation as a secondary or tertiary identity, if acknowledged at all. We changed the application so that members of the AEA could identify their primary area specialization as EVALUATION and thereby assert a primary identity as evaluator. Two decades later, in a much-changed context, it is appropriate to assert our identity as evaluation scientists engaged in evaluation science. How we identify ourselves, both individually and collectively, affects how we present ourselves to others and how they see us. How others view us has social, cultural, economic, academic, and political implications. The Thomas Theorem applies: *What is perceived as real is real in its consequences* (Thomas & Thomas, 1928, p. 572). To make common cause with other scientists, it helps if we perceive ourselves as scientists, or so it seems to me.

Conclusion

If we are to seek recognition of evaluation as a branch of science among other branches of science, we must begin by recognizing, indeed, embracing our foundation in science. If we are to seek support for funding evaluation, we might do well to make common cause with funding for all branches of science, basic and applied. If we are to play our part in combatting antiscience attitudes and actions, which are inherently also anti-evaluation attitudes and actions, then we would do well to make common cause with other scientists.

The core consequence of perceiving and asserting evaluation as science, I am positing, is that it enhances our credibility, responsibility, capability, utility, and effectiveness while communicating our role more clearly and credibly to those who value science but haven't thought of evaluation as a scientific activity. Positioning evaluation as science may also have consequences for how we are viewed, treated, perceived, and located in academic institutions, government agencies, and by funders and users of evaluation.

AEA's endorsement of the March for Science on behalf of the evaluation community was a political act. As evaluation scientists, we can be explicit and active in supporting the importance of science in our world and make common cause with other scientists in supporting and advocating for science. I do so as an evaluation scientist.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note

1. For their contributions in focusing, clarifying, and improving this article, I am deeply appreciative of the feedback from anonymous peer reviewers and the in-depth review and suggestions from the AJE editor.

References

- Alkin, M. (Ed.). (2013). *Evaluation roots* (2nd ed.). Los Angeles, CA: Sage.
- Alkin, M., & Christie, C. (2004). An evaluation theory tree. In M. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 12–65). Thousand Oaks, CA: Sage.
- American Association for the Advancement of Science. (2017). Retrieved from <https://www.aaas.org/about/mission-and-history>
- Argyris, C., Putnam, R., & Smith, D. M. (1985). *Action science*. San Francisco, CA: Jossey-Bass.
- Blackburn, S. (2005). *Truth: A guide*. New York, NY: Oxford University press.
- Bradford, A. (2015). Science & the scientific method: A definition. *LiveScience*. Retrieved from <https://www.livescience.com/20896-science-scientific-method.html>
- Bronowski, J. (1977). *A sense of the future*. Quoted in A. Lewis (Ed.). (1980), *The quotable quotations book*. New York, NY: Crowell.
- Chavis, D. (2015). *Community science*. Retrieved from <http://www.communityscience.com/>
- Christie, C. A., Inkelas, M., & Lemire, S. (Eds.). (2017). Improvement science in evaluation: Methods and uses. *New Directions for Evaluation*, 153.
- Clark, W. C. (2007). Sustainability Science: A room of its own. *Proceedings of the National Academy of Sciences*. Washington, DC: National Academy of Sciences, p. 1738.
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco, CA: Jossey-Bass.
- Cronbach, L. J., & Associates. (1980). *Toward reform of program evaluation*. San Francisco, CA: Jossey-Bass.
- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. New York, NY: Lawrence Erlbaum.
- Donaldson, S. I. (2017). *Evaluation science*. AEA eStudy webinar, June 20.
- Drucker, P. (1959). *The landmarks of tomorrow*. New York, NY: Harper and Row.
- Duffin, J. (2010). *History of medicine* (2nd ed.). Toronto, Canada: University of Toronto Press.
- Eccles, M. P., & Mittman, B. S. (2006). Welcome to implementation science. *Implementation Science*, 1. Retrieved from <https://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-1-1>
- Fleur, N. S. (2017, April 22). Scientists, feeling under siege, march against Trump policies. *The New York Times*. Retrieved from https://www.nytimes.com/2017/04/22/science/march-for-science.html?emc=edit_nn_20170426&nl=morning-briefing&nid=44499219&te=1&r=0
- Goldacre, B. (2009). *Bad science*. London, England: Fourth Estate.
- Greene, J. C., & Henry, G. (2005). Qualitative-quantitative debate in evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 345–350). Thousand Oaks, CA: Sage.
- Guttentag, M., & Struening, E. L. (1975). *Handbook of evaluation research* (Vols. 1-2). Beverly Hills, CA: Sage.
- Harris, R. F. (2017). *Rigor mortis: How sloppy science creates worthless cures, crushes hope, and wastes billions*. New York, NY: Basic Books.

- Hopson, R. (Ed.). (2000). How and why language matters in evaluation. *New Directions for Evaluation*, 86.
- Jane Fields, J., & Sheldon, T. (2017). *Aligning concepts and language demystifies evaluation*. AEA365 blog, July 9. Retrieved from http://aea365.org/blog/aligning-concepts-and-language-demystifies-evaluation-by-jane-fields-and-tim-sheldon/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+aea365+%28AEA365%29
- Lasswell, H. D. (1970). *Policy sciences*. Retrieved from <http://www.policysciences.org/>
- Lin, L., & Spector, J. M. (Eds.). (2017). *The sciences of learning and instructional design: Constructive articulation between communities*. New York, NY: Routledge.
- Madaus, G. F., Scriven, M., & Stufflebeam, D. L. (Eds.). (2012). *Evaluation models: Viewpoints on educational and human services evaluation* (Vol. 6). Amsterdam, the Netherlands: Springer.
- Management and Service Science. (2017). 11th International Conference on Management and Service Science, Guilin, China, Oct 23–25. Retrieved from http://www.engii.org/conference/MASS/?utm_campaign=mass&utm_source=e_cp&utm_medium=conf_mass_d2_20170718_pan_119747
- Mathison, S. (Ed.). (2005). *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage.
- Miner, M., & Rawson, H. (Eds.). (2000). *The new International dictionary of quotations* (3rd ed.). New York, NY: Signet.
- Morell, J. (1979). *Evaluation as social technology*. Retrieved from <https://evaluationuncertainty.com/evaluation-as-social-technology/>
- Nelson, C. A., de Haan, M., & Quinn, P. C. (Eds.). (2016). *Developmental science*. New York, NY: John Wiley.
- Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (Eds.). (2015) *Handbook of practical program evaluation* (4th ed.). Hoboken, NJ: Wiley.
- Offit, P. (2017). *Pandora's lab: Seven stories of science gone wrong*. Washington, DC: National Geographic.
- Olmstead, M. (2017). *Seeking the unheard voices of science: How science journalists consider diversity when finding sources*. Retrieved from <https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/61082/ProjectReport.pdf?sequence=1>
- Pardridge, W. M. (2003). *Translational science: What is it and why is it so important?* doi: 10.1016/S1359-6446(03)02823-X
- Partnow, E. (Ed.). (1977). *The quotable woman: An encyclopedia of useful quotations, Vol. 1, 1800–1899*. Los Angeles, CA: Pinnacle books.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Los Angeles, CA: Sage.
- Patton, M. Q. (2012). *Essentials of utilization-focused evaluation*. Los Angeles, CA: Sage.
- Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). Los Angeles, CA: Sage.
- Patton, M. Q. (2018). *Facilitating evaluation: Principles in practice*. Los Angeles, CA: Sage.
- Pawson, R. (2013). *The science of evaluation: A realist manifesto*. London, England: Sage.
- Peter, L. J. (1977). *Peters's quotations: Ideas for our time*. New York, NY: Bantam books.
- Rossi, P., Lipsey, M., & Freeman, H. (2003). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Rossi, P. H., Freeman, H. E., & Wright, S. R. (1979). *Evaluation: A systematic approach*. Beverly Hills, CA: Sage.
- Schwandt, T. (2002). *Evaluation practice reconsidered*. New York, NY: Peter Lang.
- Science. (n.d.). In *Oxford Dictionaries*. Retrieved from <https://en.oxforddictionaries.com/definition/science>
- Scriven, M. (1976). *Reasoning*. New York, NY: McGraw-Hill.
- Scriven, M. (1991). *Evaluation thesaurus*. Beverly Hills, CA: Sage.
- Scriven, M. (1993). Hard-won lessons in program evaluation. *New Directions for Program Evaluation*, 58.
- Scriven, M. (2003, Winter). Differences between evaluation and social science research. *The Evaluation Exchange*, 9, 1.
- Scriven, M. (2004). Reflections. In M. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 183–195). Thousand Oaks, CA: Sage.

- Scriven, M. (2008). The concept of a transdiscipline: And of evaluation as a transdiscipline. *Journal of Multi-disciplinary Evaluation*, 5, 65–66.
- Scriven, M. (2013). Conceptual resolutions and evaluation: Past, present, and future. In M. Alkin (Ed.), *Evaluation roots* (2nd ed., pp. 167–179). Los Angeles, CA: Sage.
- Scriven, M. (2016). Roadblocks to recognition and revolution. *American Journal of Evaluation*, 37, 27–44.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newberry Park, CA: Sage.
- Shaw, I. F., Greene, J. C., & Mark, M. M. (Eds.). (2006). *The Sage handbook of evaluation: Policies, programs and practices*. Thousand Oaks, CA: Sage.
- Spangenberg, J. H. (2011). Sustainability science: A review, an analysis and some empirical lessons. *Environmental Conservation*, 38, 275–287.
- Stein, J., & Urdang, L. (Eds.). (1966). *The random house dictionary of the english language* (unabridged ed.). New York, NY: Random House.
- Stufflebeam, D. L., & Coryn, C. L. C. (2014). *Evaluation theory, models, and applications* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Stufflebeam, D. L., Madaus, G. F., & Thomas Kellaghan, T. (Eds.). (2000). *Evaluation models: Viewpoints on educational and human services evaluation*. Boston, MA: Kluwer.
- Thomas, W. I., & Thomas, D. S. (1928). *The child in America: Behavior problems and programs*. New York, NY: Alfred A. Knopf.
- Townsend, A. (2015). *The future of urban science*. Retrieved from <http://www.citiesofdata.org/report-the-future-of-urban-science/>
- Webster's New World Dictionary. (2nd college ed., p. 1275). (1982). New York, NY: Simon & Schuster.
- Weiss, C. (2002). Lessons learned: A comment. *American Journal of Evaluation*, 23, 229–230.
- Wootton, D. (2015). *The invention of science: A new history of the scientific revolution*. New York, NY: Harper.