© *SpazioFilosofico* 2015 – ISSN: 2038-6788

Jennifer C. Greene

EVALUATION AS A SOCIO-POLITICAL INTERVENTION

Abstract

The logic of evaluation is presented and examined, resulting in the claim that evaluation (of public policies and programs) itself is an intervention in the public sphere. Evaluation accomplishes this primarily via the advancement of selected stakeholder interests and values, represented most evidently in the evaluation's purpose, audience, and key questions, which in turn, structure the quality criteria upon which judgments of program quality and effectiveness are made. Illustrative examples are offered.

In societies all around the globe, the contemporary practice of evaluation powerfully influences media headlines, political debates, and, most consequentially, actual social and economic policies and programs. Evaluation data on politically-contested programs are used, often by all sides in the debate, as fodder for public arguments about policy directions and program parameters. For programs of significant public interest or consequence (for example, an experimental program in health care subsidies for the poor), evaluation results capture national headlines. Interestingly, challenges to and debates about evaluation itself – its questions, designs, methodologies, and especially its values – rarely enter the public debate[1]. This is so, even though evaluation is far from a homogenous social practice, *and* the character of any particular evaluation study significantly shapes, even constitutes in part, the results obtained.

In this essay, I will take up the challenges of evaluation itself as an intervention in public policy and program decision making[2]. I will argue that the presence of an evaluation in the spaces occupied by a public program influences that program in particular ways, depending on the character of the evaluation. So, what is evaluated is a program-with-an-evaluation, rather than a program pursuing a course void of evaluative influences. For some evaluation approaches, the influences on the program are intentional and for others, more of an unintended and unobserved side-effect. The argument will begin with a general discussion of the logic of evaluation and how this logic constitutes an intervention, followed by a discussion of how the character of this

[1] The National Education Policy Center at the University of Colorado at Boulder (http://nepc.colorado.edu/) is a rare exception, although the public reach of the work of this center is not known.

[2] This essay concentrates on social-economic programs, mostly in the public sphere, that are designed to support people in need. Programs designed for the well-off or the wealthy, for example tax codes, are much more rarely evaluated (Datta-Grasso 1998) and their evaluations may have different kinds of influences.

intervention is connected to the evaluation approach being used. These discussions will focus on perhaps the most powerful lever of influence in evaluation, that of the values being advanced by the inquiry. In discussing evaluation's values, I will somewhat summarily assert that democratic values are the most defensible of evaluation's varied portfolio of value stances. Finally, I will illustrate the arguments made by offering examples of evaluation in action.

I offer these thoughts as a lifetime "theorist" and practitioner of program evaluation in the United States. My evaluation practice has focused largely on formal educational programs for children and youth, and has also included evaluations of non-formal educational programs (for example, summer camps or after-school programs) and a smattering of evaluations of programs in other domains.

1. *On Evaluation as a Values-Engaged Intervention*

This part of the argument first establishes the values parameters of evaluation and then turns to how these values parameters influentially serve to position evaluation itself as an intervention.

Quality criteria in evaluation as conveyers of values

Evaluation is the systematic assessment of the operations and/or outcomes of a program or policy, compared to a set of explicit or implicit standards, as a means of contributing to the improvement of the program or policy. (Weiss 1998, p. 4)

This definition of evaluation by the late Carol Weiss is commonly cited, both because it is relatively broad and inclusive *and* because it explicitly states that evaluation involves standards and thereby value judgments. The core of the evaluative enterprise, captured in this definition, is the comparison of empirical data collected about a given program (its underlying theory, context, operations, and/or outcomes) to established standards or criteria that define what constitutes program quality in that evaluation setting. Interestingly, few evaluation proposals or reports explicitly articulate and justify the quality criteria being used. Especially so, a key question becomes, where do these program standards or quality criteria come from, and just how are they determined?

In brief, standards or criteria for judging program quality are embedded in the logic of the particular evaluation approach being employed in that context. Table 1 outlines the generic logic of evaluation. This outline presents core elements of evaluation (in theory and in practice), beginning with the program/policy and evaluation contexts; followed by the evaluation's primary purposes, intended audiences, key questions, and the evaluation approach to be used[3]; and then the criteria to be used to judge program quality; the evaluation methodology; plans for communication and utilization; and finally, meta-evaluative criteria. Whether explicitly stated or not, these elements describe evaluation's basic logic. And, as presented in the next section, different constellations of

[3] Different evaluation approaches are differentially well suited to address different evaluation purposes, audiences, and key questions.

these elements describe different approaches to evaluation, and approaches gain coherence with strong and logical connections among their distinct elements.

Table 1
A Framework for Evaluative Logic

Evaluation element	**Description**
Program and context	Description of the program to be evaluated (needs, goals/objectives, activities, staffing) and of the organization(s) involved; the settings in which the program is being implemented; the program's policy context; and particular politics of relevance.
Evaluation context	The impetus for the evaluation – who wants it and why? Who are the "automatic" evaluation audiences? What politics of importance are related to the request for the evaluation?
Purpose	The reasons the evaluation is being conducted. Stated reasons include to improve the program, to contribute to policy decisions, to generate knowledge, to provide accountability for public funding, to promote organizational growth or change, to educate consumers, to empower participants, to catalyze social change. Unstated reasons include to politically support or undermine the program, to stall for time, to damage a political opponent.
Audience	Which stakeholders' needs for information and evaluation questions are being addressed in the evaluation? Stakeholders include funders, policy and other decision makers, program developers, program managers, onsite administrators, staff, program participants and their families and communities, interested citizenry, the media.
Questions	The particular questions the evaluation will address; evaluation questions are generally framed by purpose and audience, but require further specification. Sample evaluation questions for a *process evaluation* of a science education program are: • How well does the curriculum engage students of varying learning styles and achievement histories? • How well does the professional development training prepare teachers to implement the program effectively?
Approach	The evaluation approach (or approaches) to be used in the evaluation and the rationale for this approach (or approaches). Extant evaluation approaches include decision-oriented, responsive, educative, utilization-focused, accountability-oriented, democratic, and praxis-oriented evaluation.
Judging program quality	The criteria or standards to be used to make judgments of program quality, the justification for these criteria, and the process to be used for making such judgments, including who is to be involved

Design and methods	The overall methodological design to be used in the evaluation, and the specific data collection and analysis methods that will be used within this design. The criteria or standards to be used to make judgments of program quality, the justification for these criteria, and the process to be used for making such judgments, including who is to be involved.
Communication and reporting	How the evaluation team will communicate and report to key clients and other audiences, both during the evaluation and at the end of the study.
Utilization	Intended uses of the evaluation process and results. Categories of common evaluation uses are instrumental, conceptual, symbolic, and political use.
Meta-evaluation	The criteria or standards to be used to make judgments of evaluation quality, and the process for making such judgments, including who is to be involved.

Collectively, the other elements in this evaluative logic serve to determine the element representing the standards or criteria for judging program quality. Most centrally, these criteria are determined by the evaluative elements that directly engage the political and values-laden evaluative issues of whose interests are being addressed by the evaluation, namely, what is the purpose of the evaluation, who is it for, and what specific questions will the evaluation address[4]. Different combinations of these elements invoke different criteria for judging quality. For example, an evaluation conducted to inform a policy decision about whether to reauthorize and refund a given program clearly addresses the outcomes-oriented information needs of policy makers. In this context, a good or high quality program is one that meets its intended outcomes. As another example, an evaluation conducted to better understand how students from different socio-demographic groups respond to an innovative technology curriculum addresses the educational information needs of program designers, and likely teachers as well. In this context, a good program is one that serves the distinct learning profiles of students from multiple socio-demographic groups equally well.

So, the values advanced by a given evaluation are most readily apparent in the criteria developed to judge program quality, which, in turn, are most directly influenced by evaluation purpose, audience, and key questions. Specific quality criteria in evaluation are drawn from varied sources, which include: (a) stated policy goals or program objectives; (b) facets of the program's theory; (c) implicit goals and objectives, unstated assumptions; (d) relevant theory and research; (e) salient dimensions of the context (for example, political or cultural factors); and (f) key stakeholder or evaluator commitments. Again, because quality criteria themselves are typically not stated, neither are the sources from which they are drawn.

[4] Renowned American evaluation theorists Shadish, Cook, and Leviton (2001) include valuing as one of the five core components of evaluation. The other four components are knowledge, use, practice, and social programming. And international evaluation theorist Michael Scriven presents the core logic of evaluation as centered on judgments of quality based on comparisons of empirical results to established standards (http://michaelscriven.info/).

Evaluation as an intervention

So, in what ways is evaluation itself an intervention and what role do the value dimensions of evaluation play in this conceptualization of the evaluation enterprise?

Broadly, *evaluation perturbs the program and context being evaluated by infusing (some form of) evaluative thinking and perspectives into ongoing program conversations and reflections.* And a competent and thoughtful evaluator, of whatever theoretical persuasion, will do this with purpose and intention, aiming for influences that enhance the aims and objectives of the particular evaluation approach being implemented and that advance the values of his/her preferred approach. These conversations and reflections, then, afford ongoing opportunities, and in some cases internal or external pressures, for stakeholders to rethink selected program aspirations, to reconsider who should be eligible for program participation, or even to revise a troublesome component of the program as it is being implemented. In these ways, the evaluation can function as an intervention in program design, implementation, and aspiration.

While there are likely multiple specific ways in which evaluation influences the program being evaluated and its context(s), I offer the following three, all fully interrelated, and all of which directly engage evaluation's values dimensions. As elaborated in the next section, the particular character of these three evaluative influences depends on the evaluation approach being implemented. First, evaluation influences *who participates* in ongoing program conversations. Representatives of the targeted evaluation audiences are usually the stakeholders identified for ongoing consultation and conversation about the program and its evaluation. And such audiences can range from policymakers to advocacy groups to intended program beneficiaries. In some evaluations, one or more stakeholder advisory boards are convened for just this purpose. Second, evaluation influences *the substance or content* of the ongoing conversations. That is, the ongoing evaluative conversations can draw attention to particular issues of importance, including issues overlooked in the program's vision. One example here is how well the program is reaching all types of eligible beneficiaries. Another concerns the cultural and linguistic appropriateness of the program's design and implementation for diverse kinds of participants. Third, evaluation influences the *relational fabric* of the program's administration, oversight, and implementation. All evaluators interact in some ways and establish some kind of relationships with some stakeholders. The valence and character of these interactions, also viewed as the social relations of evaluation (Abma 2006), matter. They can modestly influence particular program components, as well as more significantly shape the overall ambience of the program's context of oversight and delivery, and thereby the consequent relevance and acceptance of the evaluation results.

A brief example can illustrate the evaluator's power and obligation to exert these influences responsibly. The example comes from an award-winning evaluation (Brandon, Smith, Trenholm, & Devaney 2010), commended for its high methodological quality, its high utility in influencing policy, *and* its purposeful, fair-minded engagement with stakeholder concerns and values, on all sides of a highly emotional and contested issue. This US evaluation was of four promising abstinence education programs, all designed to reduce teen pregnancy by teaching youth about the benefits of abstinence

("just say no") and the risks of sexual activity before marriage. Politically, abstinence is promoted by conservatives while liberals favor birth control alternatives or abortion. At times, this debate can get very heated, as beliefs about these issues run deep.

The evaluation, funded by the US government, was intended to address policymakers' questions about the impact of abstinence education on sexual activity to aid future policy making. As preferred by the policy audience, the evaluators designed an experimental impact study to assess the effects of the four programs on key behavioral outcomes, including sexual abstinence, risks of pregnancy, and incidence of sexually-transmitted diseases. Survey data were collected in four waves over a 10-year period. The results found no evidence that abstinence education had increased rates of sexual abstinence or in any way affected sexual risks among participating youth.

In the context of the present argument about evaluation as a values-engaged intervention, what was most significant about this evaluation was how the evaluation team intentionally worked to foster ongoing discussion among holders of competing values and standpoints in this highly charged political context of abstinence education, keeping both detractors and advocates respectfully engaged in the evaluation throughout the long 10-year period. The evaluators chose not to ignore the politics of the evaluation – as is customary in experimental work – but rather to engage them head-on. Specifically, the evaluators inclusively chose to allocate time and resources to listening and dialoguing with stakeholders who held opposing views on abstinence. The evaluators established a national technical advisory group, comprised of technical experts with diverse stances on abstinence, and consulted with them on multiple technical issues throughout the evaluation. And they used a parallel process to engage diverse program stakeholders at each local level, again demonstrating respect for program staff. In a report on the evaluation process the evaluators said, "You need to understand the program from the local stakeholders' perspective and why they think the program is worthy and how they implement it. And you need to get their buy-in so they believe the study is credible."

In this study, the evaluators primarily addressed the first evaluative influence presented above, that of fostering reflective and critical conversation among an intentionally diverse set of stakeholders, instead of just with designated policymakers and/or key program staff. This ongoing conversation influenced some measurement decisions, enhanced participant understanding of the specific nature of the four abstinence programs, and significantly increased the credibility of the evaluation results. The evaluation, that is, constituted a substantial intervention in the policy and program context that surrounded the abstinence debate at that time.

2. *The Multiple Countenances of Evaluation as an Intervention*

In this section, I will present brief intervention-and-values profiles for two contrasting approaches to evaluation from among five broad families of evaluation approaches[5].

[5] These families are differentiated by their major foci: (1) outcomes and policy, (2) performance and accountability, (3) understanding and learning about the social problem at hand; (4) deep contextual

These profiles follow the general evaluation logic of Table 1, and are intended to further illustrate the argument made above regarding the values-based character of evaluation as an intervention. The two approaches are (a) the outcomes and policy-oriented evaluation approach, arguably the most common evaluation in our public sectors, and (b) the democratic evaluation approach, which most explicitly promotes particular values – specifically, equity, fairness, and justice – throughout its practical enactment. The context for both profiles will be an evaluation of an innovative web-based curricula and resources for students and teachers in middle school science (for children ages 11-14). The program, *Science for All,* has been thoroughly field-tested and revised and is now being implemented on a pilot basis for two years in three school districts (urban, suburban, and rural) in one state. The evaluation is being conducted by a well-established evaluation center at a state university. Both the program and the evaluation are funded by state education dollars.

Policy-oriented evaluation: Intervention as "business as usual"

The policy-oriented evaluation team would view state education policy makers as the primary audience for the evaluation, as they are the funders and decision makers in this context. Likely decisions resulting from this evaluation (which also constitute the primary intended evaluation uses) include continuation of the program and possible expansion to other schools and districts. Secondary audiences would be the administrators and teachers in the pilot schools. The team would view the main evaluation purpose as assessing how well the program "works" or how well it accomplishes its intended outcomes of science learning. Though not likely stated as such, attainment of intended outcomes would also constitute the primary criteria for judging program quality and effectiveness. In turn, key evaluation questions would focus on outcome attainment, likely in terms of average scores. Resources permitting, evaluation questions about teacher (and student) experiences with the *Science for All* program may also be posed. Consonant with these evaluative decisions, an experimental or quasi-experimental methodology would likely be selected, as the experimental counterfactual (what happens in similar contexts without the program) can provide strong evidence on the attribution of observed outcomes to the program being evaluated. An advisory board of science education and experimental evaluation experts may be convened. The evaluators would likely provide regular progress reports (written and oral) to this board and to the key evaluation audiences.

In terms of the three evaluation-as-intervention influences on the program and its context discussed above, in this policy-oriented evaluation, (1) participants in evaluative conversations about the program are primarily existing decision makers, (2) these conversations are focused on the program as designed and especially intended outcomes, and (3) the relationships established via the evaluation do not likely challenge extant hierarchies, boundaries, or norms. Therefore, the values advanced in this evaluation include support for the decision making status quo, a valuing of ends (learning outcomes) over means (learning experiences), a utilitarian emphasis on average effects,

---

and socio-cultural understanding of the problem at hand; and (5) democratization and socio-political critique.

and methodological objectivity and distance. This evaluation as intervention would not disturb the status quo.

Democratic evaluation: Intervention as disturbance or disquiet

The democratic evaluation team would design an evaluation study and communicative process that are significantly different from the familiar policy-oriented evaluation described just above. Following the democratic ideas of Ernest House (House 2014; House-Howe 1999), these evaluators would seek to equitably include representatives from *all* important stakeholder groups as key audiences for the evaluation – from decision makers and administrators to science education experts and teachers, and also to students participating in the program and their parents, along with relevant community and media representatives. All audiences would be consulted regarding their priorities for evaluation questions, and respectful dialogues among the various stakeholder audiences would be encouraged as vehicles for learning about each other's program standpoints and perspectives. Advisory boards comprising representative stakeholders may be established. The evaluation's purposes would include both substantive learning about the quality and effectiveness of the *Science for All* program, as disaggregated by relevant student sub-groups, and advancement of democratic values of equity, inclusion, and social justice in the contexts at hand. Relatedly, key evaluation questions would ask how well the program serves students from various socio-cultural and demographic groups, and in particular, how well the program serves students who are least well served in the relevant contexts. Key evaluation questions would also engage both the quality of the learning experience and the magnitude and contextual importance of the learning outcomes. And criteria for judging program quality would focus on equity and fairness for all students, in addition to educational soundness in access to program opportunities, experiences, and outcomes. A variety of methodologies are likely to be employed, to gather data relevant to the full range of evaluation questions. And the evaluation team would communicate and consult with stakeholder groups throughout the evaluation process (orally and in writing), keeping stakeholders informed of the evaluation's progress and seeking their input on key evaluation decisions.

In terms of the three evaluation-as-intervention influences on the program and its context discussed above, in this democratic evaluation, (1) participants in evaluative conversations about the program span the full range of program stakeholders; (2) these conversations are focused on the programmatic and educational interests and concerns of various stakeholders; and (3) the relationships established in the evaluation aspire to be respectful and dialogic. Therefore, the values advanced in this evaluation include inclusion, equity, and educational quality in program access, experience, and outcome for all students, especially those under-served. This evaluation as intervention would likely disturb the status quo, in ways necessary for our societies to reach their full democratic ideals.

3. *Reprise*

Not all readers will agree with my support for a democratic approach to evaluation, as is expected and desired in our pluralistic societies. Even so, I still encourage all readers to recognize the interventionist strands of their own work as evaluators; to reconfigure these strands as necessary to enact a defensible evaluation practice (from each reader's point of view); and also to name, claim, and justify the values that are advanced by our respective evaluation practices.

References

- T.A. ABMA (2006), *The Social Relations of Evaluation*, in I.F. SHAW-J.C. GREENE-M.M. MARK (eds.), *The Sage Handbook of Evaluation*, Sage, London 2006, pp. 184-199.
- P.R. BRANDON-N.L. SMITH-C. TRENHOLM-B. DEVANEY (2010), *Evaluation Exemplar: The Critical Importance of Stakeholder Relations in a National, Experimental Abstinence Education Evaluation*, in "American Journal of Evaluation", 31 (4/2010), pp. 517-531.
- L.-E. DATTA-P.G. GRASSO (eds.) (1998), *Evaluating Tax Expenditures: Tools and Techniques for Assessing Outcomes* ("New Directions for Evaluation", 79 (1998)).
- E.R. HOUSE (2014), *Evaluating: Values, Biases, and Practical Wisdom*, Information Age Publishing, Charlotte NC 2014.
- E.R. HOUSE-K. HOWE (1999), *Values in Evaluation and Social Research*, Sage, Thousand Oaks CA 1999.
- W.R. SHADISH-T.D. COOK-L.C. LEVITON (2001), *Foundations of Program Evaluation: Theories of Practice*, Sage, Thousand Oaks CA 2001.
- C.H. WEISS (1998), *Evaluation*, Prentice Hall Inc., Upper Saddle River NJ 1998[2].