

Discussant Commentary

Commentary on Generative Evaluation

Michael Quinn Patton

Utilization-Focused Evaluation
mquinnp@gmail.com

Abstract

As the field of program evaluation has grown over the last six decades, new types of evaluation have emerged to meet different needs and serve diverse and distinct purposes. This article elucidates the important evaluation niche and contribution of Generative Evaluation

Keywords

evaluation, developmental, generative, utilization-focused

I appreciate the opportunity to share some thoughts about *Generative Evaluation*. To put my commentary in context, perhaps a bit of history will be helpful. I have been engaged in evaluation full-time for six decades, most of that time as an independent consultant. That is, I make my living as an evaluation professional. I have written eight books on evaluation focused on making evaluation useful, qualitative methods applied to evaluation, how to be both practical and creative in conducting evaluations, taking a global perspective, and

incorporating systems thinking and complexity theory into evaluation designs. As a former president of the American Evaluation Association and a founding member of the International Evaluation Academy, I am committed to supporting the development of the evaluation profession worldwide.

I have witnessed the emergence of multiple distinct approaches to evaluation. As the field of evaluation has grown over the last six decades, new types of evaluation have emerged to meet different needs and serve diverse and distinct purposes. The options chapter in the 4th edition of my book on *Utilization-Focused Evaluation* catalogued 80 different ways of focusing evaluations (Patton, 2008, pp. 300–305), from A to Z: accountability, accreditation, and appreciative inquiry through process evaluation and product evaluation to theory-of-change approaches and zero-based budgeting. There are goal-based versus goal-free evaluations; both process and outcomes evaluations; participatory versus connoisseurship approaches; real-world and theory-driven frameworks; internal versus external options; cost-benefit and transformative models; constructivist and realist evaluations. Engaging in evaluation requires distinguishing types, understanding options, and choosing among alternative approaches to evaluation that serve the distinct needs of those who will use the evaluation in a given context for a specific purpose. A recent video inventory listing distinct evaluation approaches identified 101 distinct types of evaluation—then stopped counting (Patton, 2022a). That video just provided a panoramic view of the great diversity of evaluation approaches but raised questions from viewers about why there are so many approaches. That led to a series of videos explicating evaluation’s diversity:

- Why so damn many options? The 10 competing values that explain the panoramic evaluation landscape. (Patton, 2022b)
- From Evaluation Theory Tree to Theory Forest: A Framework for Depicting Alternative Evaluation Approaches (Patton, [2023a](#))
- The Periodic Table of Evaluation (Patton, 2024b)
- Evaluation’s Story in Fable and Song: Greatest Hits Over Five Decades (Patton, 2023b)
- Honoring the 7 Directions: An Indigenous Framework for Designing and Evaluating Systems Change (Patton, 2024a)

As this collection of videos illustrates, evaluation is characterized by diversity. As different approaches have emerged, each gets a name, a brand, a label, a way of distinguishing it from other approaches. *Generative Evaluation* follows that tradition constituting a new distinction as conceptualized by Malika Virah-Sawmy, Christina Tewes-Gradl, and Pierre Golbach (2024) in the article to which this is a response. Where, then, does *Generative Evaluation* fit in the diverse evaluation landscape and what is its unique contribution? In part because of my work in tracking and cataloging diverse and emergent approaches to evaluation, Malika Virah-Sawmy contacted me to have a dialogue about

Generative Evaluation. That conversation is captured in two videos we did together about the distinctive niche of *Generative Evaluation* (Patton & Virah-Sawmy, 2024a) and *Unpacking System Change* (Patton & Virah-Sawmy, 2024b). What Malika and I discussed in those dialogues is extended in this commentary. First, however, I will share a bit more about what I see as the larger evaluation context to position and elucidate the important evaluation niche and contribution of *Generative Evaluation*.

Distinctions matter. Language matters. Terminology matters. One way of keeping track of developments in the field of evaluation is the field's major topical journal, *New Directions for Evaluation*, published quarterly by the American Evaluation Association. A special volume of the journal was devoted to *How and Why Language Matters in Evaluation* (Hopson, 2000). In my article in that volume (Patton, 2000), I opened by recounting a classic story from that great scholar and observer of all things human, Dr. Seuss (1961), a story that illustrates the consequences of not making distinctions. His children's story of *Too Many Daves* tells us that Mrs. McCave had 23 sons, and she named them all Dave. When she wanted one particular Dave, and called out his name, all 23 Daves came on the run. The same thing happens if you don't distinguish types of evaluations. You cannot just do an evaluation. You have to do a particular kind of evaluation.

From a Command-and-Control Project Mentality to Living Systems Complexities and Uncertainties

While evaluative thinking, inquiry, and judgments are as old as and can be thought of as inherent to our human species, formal and systematic evaluation as a field of professional practice is relatively recent. Evaluation in the United States grew up in the "projects"—federal government projects spawned by the Great Society legislation of the 1960s. The "War on Poverty" initiated by President Lyndon Johnson led to massive federal expenditures on a great variety of programs during the 1960s and 1970s. That programming gave rise to calls for accountability. This meant more than financial audits. Demand for systematic empirical evaluation of the effectiveness of government programs grew as government programs grew. That demand for accountability spread to philanthropic foundation initiatives, international development agencies, and nongovernmental organizations worldwide. Those wide-ranging development initiatives shared a common logic. Evaluation for accountability was, and still is, embedded in that logic.

The traditional logic of programming is a command-and-control mentality: *plan your work, work your plan, and evaluate the results*. Programs are planned to meet some need and achieve specific, predetermined goals. Program plans then have to be implemented. Traditional evaluation determines to what extent and in what ways the program works to bring about desired changes. Do students learn what they are supposed to learn? Do farmers adopt new practices

offered by agricultural extension programs? Do people in a community recycle appropriately when recycling services are offered? It all sounds pretty straightforward. Except, as the classic proverb cautions:

The best laid plans of mice and humans often go awry.

Plans often must be adapted to the realities of a turbulent and uncertain world. A famous military adage is that *no plan survives first contact with the enemy*. This doesn't mean that planning is useless. As former US President Dwight D. Eisenhower (1957) observed, "Plans are worthless, but planning is essential." The point is to be ready to adapt plans to fit changing circumstances rather than rigidly implementing a plan without regard to changing conditions. *Generative Evaluation*, as defined in the volume this commentary accompanies, is built on a mindset of readiness for adaptation based on ongoing situation analysis, continuous evaluative questioning, responsive adaptation based on learning, and acute attention to emergence in complex dynamic living systems.

Two Contrasting Visions of Evaluation

Donald Campbell (1998), a pioneering evaluation thought leader, envisioned evaluation as the foundation for an *Experimenting Society*, a society that would vigorously try out solutions to problems, then evaluate those solutions rigorously, determining what works and what does not through ongoing innovation. *Generative Evaluation*, I would suggest, has inherited, builds on, and strives to contribute to that vision.

The problem is that the vision of an *Experimenting Society* has been undercut and overwhelmed by a contrary approach to evaluation characterized as the *Audit Society* (Power, 2011)—standardizing, routinizing, and mandating evaluation processes and procedures to ensure bureaucratic compliance and accountability. The *Audit Society* aims to control the world through standardized procedures, policies, and checklists that operationalize systemic accountability. The *Audit Society* strives to minimize risk and maximize certainty in the face of an increasingly chaotic and unpredictable world. The *Audit Society* establishes standardized protocols and mandated reporting mechanisms to maintain control and emphasizes accountability through paper trails and compliance monitoring. The *Audit Society* values standardization over customization, judgment over learning, consistency over innovation, and predictability over adaptation.

Judgment-oriented (summative) evaluation requires clear, specific, and measurable outcomes that are to be achieved through processes detailed in a linear logic model. This demand for upfront, preordained specificity does not work under conditions of high innovation, exploration, uncertainty, turbulence, and emergence. Indeed, premature specificity can do harm and generate resistance from social innovators, as, indeed, it has, by constraining exploration, limiting adaptation, reducing experimental options, and forcing premature adoption of a rigid model, not because such a model is appropriate, but because evaluators, funders, or other stakeholders demand it in order to comply with

what they understand to be good evaluation. *Generative Evaluation* offers an important alternative to standardized compliance-focused evaluation.

Purpose-Driven

Generative Evaluation is purpose-driven. The authors write:

Generative Evaluation helps us learn about how well the design serves the intended purpose in real life; to what extent the purpose is actually shared by various actors; and what the living system teaches us about adjusting the purpose and design. (Virah-Sawmy, Tewes-Gradl, & Golbach, 2024, p. 136)

Purpose-driven learning, grounded in relationships, is inherently emergent and generative. It follows these premises:

- Purpose drives both program operations and execution—as well as evaluation responses.
- Different kinds of evaluation serve different purposes and occupy different niches in the world of programming and change initiatives. To contribute significantly within a niche requires understanding and serving its purpose.
- Generative Evaluation serves a distinct purpose in the broad and diverse evaluation landscape.

From my perspective, the language of *Generative Evaluation* suggests a purpose for evaluation that goes beyond simply looking at what was done and judging what was done through the traditional accountability lens. Rather, *Generative Evaluation* makes evaluative thinking and dialogue part of the ongoing, complex, dynamic, and generative process of program and systems change development. This involves bringing evaluative thinking and questions into the creation, innovation, and adaptation processes and connecting individual actors to systems change dynamics using, for example, *Theory U* (Scharmer, 2018). Individuals change as part of the generative evaluation process. Relationships are formed and nurtured, which are themselves generative and, in being so, affect the ongoing development process.

One of the controversies about *Developmental Evaluation* (Patton, 2011) is that the evaluator becomes involved in and has an impact on programming. The traditional Audit Society view is that evaluators are supposed to be external, independent, objective, and neutral which, in my experience, renders them clueless about what is happening developmentally. In contrast, *Generative Evaluation* makes it clear that evaluation can and should contribute to a generative process where everyone is transformed as part of transforming systems. That is pretty scary for traditional evaluators. But it is a welcome opportunity and positioning for evaluators who want to contribute more than just writing accountability reports.

The niche and contributions of *Generative Evaluation* are well-articulated in this article. The processes of dialogic co-creation based on systemic inquiry within a living systems framework is especially relevant and appropriate in a world where transforming systems to create a more sustainable and equitable future must be the overarching purpose of both programming and evaluation. That purpose will be well-served in striving to realize the vision of an experimenting society based on shared learning and collective action informed by truly *Generative Evaluation*.

References

- Campbell, D. T. (1998). *The experimenting society*. In W. N. Dunn (Ed.), *The experimenting society: Essays in honor of Donald T. Campbell* (pp. 35–68). Transaction Publishers. <https://doi.org/10.4324/9781003578192-2>
- Dahler-Larsen, P. (2011). *The evaluation society*. Stanford University Press. <https://doi.org/10.2307/j.ctvqsdq12>
- Eisenhower, D.D. (1957). *Remarks at the National Defense Executive Reserve Conference, 11/14/57*. [Quotes | Eisenhower Presidential Library](https://www.eisenhowerlibrary.gov/record/111457).
- Hopson, R. K. (Ed.) (2000). *How and why language matters in evaluation*. *New Directions for Evaluation*, 85. <https://doi.org/10.1002/ev.1167>
- Patton, M. Q. (2000). Language matters. *New Directions for Evaluation*, 86, 5–16. <https://doi.org/10.1002/ev.1168>
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Sage Publications.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance evaluation and use*. Guilford Press.
- Patton, M. Q. (2022a, May 16). *Panorama of evaluation approaches* [Video]. YouTube. Retrieved from <https://youtu.be/GEGtBnkDyBk>
- Patton, M. Q. (2022b, June 13). *Why so damn many options? The 10 competing values that explain the panoramic evaluation landscape* [Video]. YouTube. Retrieved from https://youtu.be/YReAFxv_31s
- Patton, M. Q. (2023a, June 14). *From evaluation theory tree to theory forest: A framework for depicting alternative evaluation approaches* [Video]. YouTube. Retrieved from <https://youtu.be/pqBJxm6Dn7k>
- Patton, M. Q. (2023b, December 20). *Evaluation's story in fable and song: Greatest hits over five decades* [Video]. YouTube. Retrieved from <https://youtu.be/ERjuQBLSRQo>
- Patton, M. Q. (2024a, August 9). *Honoring the 7 directions: An indigenous framework for designing and evaluating systems change* [Video]. YouTube. Retrieved from https://youtu.be/H_vDhZq-4uk
- Patton, M. Q. (2024b, August 22). *The periodic table of evaluation* [Video]. YouTube. Retrieved from <https://youtu.be/MxIbwlbQIKI>
- Patton, M. Q., & Virah-Sawmy, M. (2024a, November 5). *Generative and developmental evaluation*. YouTube. Retrieved from <https://youtu.be/F6mgdIt79lA>
- Patton, M. Q., & Virah-Sawmy, M. (2024b, November 5). *Unpacking system change*. YouTube. Retrieved from <https://youtu.be/985FuZcZ910>
- Power, M. (2011) *The audit society: rituals of verification*. Oxford Academic.

- Seuss, D. (1961). *The Sneetches and other stories*. Random House.
- Scharmer, C.O. (2018). *The essentials of Theory U: Core principles and applications*. Berrett-Koehler Publishers.
- Virah-Sawmy, M., Tewes-Gradl, C. & Golbach, P. (2024). Generative Evaluation: Learning with and for living systems. *Journal of Awareness-Based Systems Change*, 4(2), pp. 131–150.