What is Essential in Developmental Evaluation? On Integrity, Fidelity, Adultery, Abstinence, Impotence, Long-Term Commitment, Integrity, and Sensitivity in Implementing Evaluation Models

Michael Quinn Patton[1]

Abstract

Fidelity concerns the extent to which a specific evaluation sufficiently incorporates the core characteristics of the overall approach to justify labeling that evaluation by its designated name. Fidelity has traditionally meant implementing a model in exactly the same way each time following the prescribed steps and procedures. The essential principles of developmental evaluation (DE), in contrast, provide high-inference sensitizing guidance that must be interpreted and applied contextually. In lieu of operationalizing DE fidelity criteria, I suggest addressing the *degree of manifest sensitivity* to essential principles. Principles as sensitizing concepts replace operational rules. This means that sensitivity to essential DE principles should be explicitly and contextually manifest in both processes and outcomes, in both design and use of findings. Eight essential principles of DE are identified and explained. Finally, 10 threats to evaluation model fidelity and/or degree of manifest sensitivity are identified with ways to mitigate those threats.

Keywords

developmental evaluation, fidelity, principles, sensitizing concepts

Bob Williams was awarded the 2014 AEA Lazarsfeld Theory Award for his cumulative contributions bringing systems approaches into evaluation. In accepting at the Awards Luncheon in Denver, attended by more than 400 evaluators, Bob explained the origin of the word system.

[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

> The word 'system' comes from the Greek word *synhistonai* (a verb incidentally, not a noun) meaning 'to stand together'. So I'd like to invite all that can do so to stand for a moment.

> I'm now going to ask some of you to sit down. I'd like to remain standing anyone who to some extent feels that you have applied systems and complexity ideas—of whatever shape or form and to a greater or lesser extent—in your evaluation practice.

Almost no one sat down. One person among the distinguished guests at the head table on stage sat, noticed that all others remained standing, and quickly stood back up.

Bob was visibly flabbergasted and said, "Well, that's not what I expected."

In the Systems in Evaluation Topical Interest Group session that followed the luncheon, some questions arose: What understandings and perceptions of applying systems approaches in evaluation had kept that diverse group of people on their feet? Do these people share common understandings and perceptions? Has systems thinking really permeated so deeply into evaluators' practice? Do their practices actually incorporate the core characteristics of the overall approach? Thus, did the fidelity problem arise for those who view systems thinking as central to their practice.

The questions about application of systems thinking in evaluation could be applied to almost any approach. This article will lay out the scope of the *fidelity challenge* in implementing distinct evaluation approaches, illustrate the challenge with specific examples, and use developmental evaluation (DE) to introduce a new way of thinking about and dealing with fidelity. I'll also examine some of the threats to fidelity and ways of handling those threats.

The Fidelity Challenge in Implementing Distinct Evaluation Approaches

An experienced DE practitioner recently told me: "More often than not, I find, people say they are doing Developmental Evaluation, but they are not."

The fidelity challenge concerns the extent to which a specific evaluation sufficiently incorporates the core characteristics of the overall approach to justify labeling that evaluation by its designated name. Just as fidelity is a central issue in efforts to replicate effective programs to new localities (are the replications faithful to the original model on which they are based?), evaluation fidelity concerns whether an evaluator following a particular model is faithful in implementing all the core steps, principles, and processes of that model.

- What must be included in a "theory-driven evaluation" to justify its designation as *theory-driven* (Coryn, Noakes, Westine, & Schröter, 2011)?
- What is the core of appreciative inquiry evaluation (MacCoy, 2014; Preskill & Catsambas, 2006)?
- What are the essential elements of Aboriginal Program evaluations (Jacob & Desautels, 2014) or indigenous methods generally (Chilisa, 2012)?
- What must occur in a participatory evaluation to deem it *genuinely* participatory (Cousins, Whitmore, & Shulha, 2014; Daigneault & Jacob, 2009)? Cousins and Chouinard (2012) reviewed 121 pieces of empirical research on participatory evaluation published from 1997 through 2011 and found great variation in approaches conducted under the "participatory" umbrella.
- What is essential to utilization-focused evaluation (Patton, 2008, 2012)? I've seen a great many evaluations labeled *utilization-focused* that provided no evidence that primary intended users had been identified and engaged to focus the evaluation on those users' priorities.
- And what, you may wonder, are the core systems ideas that Bob Williams was referring to in the incident above? He coedited the first expert anthology on *Systems Concepts in Evaluation*

(Williams & Iman, 2007). He estimates that as many as a thousand separate frameworks and methods fall under the systems banner because there is no single agreed-upon definition of a system. Nor is there ever likely to be. But having analyzed the great variety of systems frameworks, he asserts that there are three core ideas that characterize systems thinking: attention to the inevitable arbitrariness of boundaries, taking into account a variety of perspectives, and dynamic, entangled inter-relationships. This means that a systems approach to evaluation must specify how boundaries of the evaluation were determined and the implications of those boundaries; whose perspectives are included (and whose omitted), again with what implications; and that inter-relationships must be documented and analyzed (Williams, 2005, 2008, 2014; Williams & Hummelbrunner, 2011; Williams & Iman, 2007; Williams & van 't Hof, 2014).

Fidelity to the Formative–Summative Distinction

To illustrate the challenge of fidelity, consider our oldest, most basic, and most sacrosanct distinctions: formative and summative. The formative–summative distinction was first conceptualized for school curriculum evaluation by philosopher and *evaluator extraordinaire* Michael Scriven (1967). He called evaluating a curriculum to determine whether it should be approved and disseminated for widespread adoption of a *summative evaluation*, evoking a summit-like decision or a summing-up of effectiveness. Summative evaluation began as a purpose designation, the purpose being to inform a major decision about the merit, worth, and significance of a curriculum, program, product, or other intervention to determine its future (kill it, cut it back, continue as is, enlarge it, and take it to scale). But the term summative evaluation quickly expanded to designate any evaluation conducted at the end of a program or project. A distinct and important purpose morphed into a timing designation: end of a program. Which is ironic—and distorting—since most summative decisions must be made months before the actual end of a program, long before evaluation final summative reports are submitted. I review scores of reports labeled "summative" every year, virtually none of which are written or timed in such a way as to inform an actual summative decision by identifiable summative decision-makers. Fidelity to the original meaning of summative has been largely lost from my perspective.

And what of formative evaluation? Scriven argued wisely that before a curriculum was summatively evaluated, it should go through a period of revision and improvement, working out bugs and problems, filling in gaps, and getting student reactions, to ensure that the curriculum was ready for rigorous summative testing. The purpose of formative evaluation was to form, shape, standardize, and finalize a model so that it was ready for summative evaluation. But, as happened with summative evaluation, the idea of formative evaluation morphed from its original purpose as the label came to be applied to any evaluation that improves a program. Though formative and summative designations were conceptualized hand in glove, the purpose of formative being to get ready for summative, that expectation often goes unfulfilled. Just as summative morphed from a clear purpose to a matter of timing (end of program), I now see a great many midterm evaluations designated as "formative" simply because the evaluation takes place in the middle of a funding cycle. These supposedly "formative" evaluations are often midterm accountability exercises determining whether the program is adhering to implementation specifications and meeting milestone performance measures. From my perspective, fidelity to the original meaning of formative has been largely lost. For example, in teaching evaluation workshops, I regularly find participants equating formative evaluation with process evaluation and summative evaluation with outcomes evaluation. Not so.

Let me reiterate. Scriven originated the formative–summative distinction under the assumption that the purpose of evaluation is *to test and judge a model*. Formative evaluations were meant to

improve the model. Summative evaluations were meant to test the model and judge its merit, worth, and significance based on whether it produces the desired outcomes and those outcomes can be attributed to the program. The terms formative and summative have become dominant both within evaluation and among those who fund and use it. But evaluation practitioners have become sloppy about what actually constitutes a formative or summative evaluation and the connection between the two. The relationship has evolved from one with a clear division of labor centered around well-specified commitments, expectations, and criteria of excellence to an open noncommittal relationship: sometimes living together, sometimes not, sometimes connected, sometimes not, and sometimes true to original ideals about what the relationship was supposed to be, but more often not.

Emergence of DE

DE emerged from my commitment to respect the fidelity of formative and summative evaluation. I had a 5-year contract to evaluate a philanthropic foundation leadership program, and 2.5 years were to be formative, to stabilize and standardize the model, followed by 2.5 years to test and judge the model's effectiveness. During the formative period, the senior program staff and foundation leadership came to realize that they didn't want to create a standardized model. Instead, they realized that they would need to be continuously adapting the leadership program as the world changed. To keep a leadership development program relevant and meaningful, they concluded, they would need, over time, to continuously update and adapt what they did; who and how they recruited people into the program; use of new technologies; and being attentive to and incorporating developments in public policy, economic changes, demographic transitions, and social–cultural shifts. They came to understand that they didn't want to improve a model or test a model or promulgate a model. Instead, they wanted to keep developing and adapting the program. They wanted an approach that would support ongoing adaptation and timely decisions about what to change, expand, close out, or further develop. This was different from formative evaluation. And they concluded that they would never commission a summative evaluation because they wouldn't have a standardized model that could be summatively evaluated. Our discussions about what they wanted and needed kept coalescing around ongoing adaptation and development so we called the approach DE. (For more details about this designation and how the DE terminology emerged, see Patton, 2011, pp. 2–4.)

Let me say a bit more about the distinct purpose and niche of DE as it has developed, which will set the stage for the emergent fidelity problem in the practice of DE, and the solution I offer, for the first time, in this article.

The Niche and Purpose of DE

DE provides evaluative information and feedback to social innovators to inform adaptive *development* in complex dynamic environments. DE brings to innovation and adaptation the processes of asking evaluative questions, applying evaluation logic, and gathering and reporting evaluative data to support project, program, initiative, product, and/or organizational development with timely feedback.

The DE niche focuses on evaluating innovations in complex dynamic environments because that's the arena in which *social innovators* are working. These are people who want to change the way things are in major ways. Innovation as used here is a broad framing that includes creating new approaches to intractable problems, ongoing program adaptation to changed conditions, adapting effective principles to new contexts (scaling), systems change, and rapid response adaptation under crisis conditions. *Social innovation* is shorthand for any kind of emergent/creative/adaptive interventions for complex problems.

Traditional evaluation approaches advocate clear, specific, and measureable outcomes that are to be achieved through processes detailed in a linear logic model. Such traditional evaluation demand

for up-front, preordinate specificity doesn't work under conditions of high innovation, exploration, uncertainty, turbulence, rapid change, and emergence. Indeed, premature specificity can do harm 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 demand it in order to do what they understand to be good evaluation. DE emerged as a response to the need for an alternative way to engage in evaluation of social innovations and adaptive developmental processes in complex dynamic environments.

The book on DE was published in 2011. In the relatively short time since, DE has become recognized and established as a distinct and useful evaluation approach (Dickson & Saunders, 2014; FSG, 2014; Lam & Shulha, 2014; Preskill & Beer, 2012). The attention garnered has also raised a major question: *What are the essential principles of DE?*

Treating Essential Principles as Sensitizing Concepts

Before listing the essential principles of DE, let me describe the developmental approach used to identify them. A core group of DE practitioners shared ideas and reactions in an interactive, clarifying, and developmental process[1]. We wanted to avoid a recipe-like or checklist approach to DE based on operationalizing key concepts and dimensions. Instead, we view these essential principles as *sensitizing concepts* that must be explicitly addressed in DE, but how and the extent to which they are addressed depends on situation and context. This is a critical departure from the usual approach to ''fidelity,'' which has traditionally meant to implement an approach operationally in exactly the same way each time. Fidelity has meant adherence to a recipe or highly prescriptive set of steps and procedures. The essential principles of DE, in contrast, provide guidance that must be interpreted and applied contextually—but *must* be applied in some way and to some extent if the evaluation is to be considered genuinely and fully developmental.

In lieu of operationalizing DE fidelity criteria, I am designating this approach: *assessing the degree of manifest sensitivity*. In lieu of fidelity, I prefer to examine the *integrity* of an approach. For a DE to have integrity, the essential DE principles should be explicitly and contextually manifest in both processes and outcomes, in both design and use of findings. Thus, when I read a DE report, talk with those involved in a DE, or listen to a DE presentation at a conference, I should be able to see/ detect/understand how these essential principles of DE informed what was done and what resulted.

Let me elaborate just a bit. The notion of judging the integrity of an approach by assessing the degree of manifest sensitivity to essential principles flows from the notion of fieldwork guided by sensitizing concepts (Patton, 2015a). A sensitizing concept raises consciousness about something, alerts us to watch out for its relevance, and reminds us to engage with the concept throughout our fieldwork within a specific context. Essential principles of DE sensitize us to what to include in DE practice.

Consider the concept *innovation*. DE is innovation-focused, one of the essential principles I'll elaborate in a moment. Here is what the concept of innovation does in a DE process. It focuses our attention on social innovators, that is, people who are trying to bring about major change. We are alerted by their definition of what they are doing (''innovation'') to find out what they mean. We pay attention to and document what they are doing and how they talk about what they are doing. We interact with them about what is going on and the implications of their efforts and documented results. We gather data about what is unfolding and emerging. We observe and provide feedback about how what is actually happening matches expectations and hopes. We work with those involved to interpret what is happening and judge what is working and not working and thereby adapt, learn, and move forward. In so doing, we are engaging with them around the notion of ''innovation'' and deepening both their and our understanding of what is meant by innovation in that context. The definition and meaning of innovation is likely to evolve, deepen, and even morph as part of the DE inquiry.

In this process, DE becomes part of the change process itself, part of the intervention. It happens like this: In inquiring into the meaning of innovation within a context and particular change-focused initiative, and providing feedback about what is learned as well as further questions generated, DE affects and alters the innovation process and outcomes.

Now then, in hearing about such an inquiry, I can make a judgment about the degree of manifest sensitivity to innovation in that particular DE and judge the integrity of the overall DE by making a similar assessment of each DE principle. Table 1 presents the eight essential DE principles (Patton, 2016).

DE as an Integrated Approach

For an evaluation to merit the label "DE," all of the principles in Table 1 should be addressed to some extent and in some way. As noted in Table 1, this is not a pick-and-choose list. All are essential. This means that there is evidence in the DE process and results that these essential principles have been addressed in some meaningful way or, for specific contextual reasons, not incorporated explicitly. For example, let's imagine working with a social innovator and/or funder who hates the word "complexity," thinks it is overused jargon, so the DE process avoids explicitly using the term *complexity* but does explicitly address emergence, adaptation, and nonlinearity. Such negotiations are part of contextual sensitivity and adaptability and part of the essential DE learning process and should be reported.

Moreover, the essential principles are interrelated and mutually reinforcing. Being utilization-focused (#1) requires staying attuned to the priority purpose of the evaluation, namely developmental (#8) support of innovation and adaptation (#2). DE occurs in a complex dynamic system (#3) that requires understanding and applying complexity concepts and systems thinking (#4) with timely feedback (#6). Utilization-focused engagement is collaborative (#5) and evaluative (#7), making the DE part of the intervention.

Empowerment Evaluation (EE) as a Contrary Example, or Not

What must be included in an EE to justify the label *empowerment*? Miller and Campbell (2006) systematically examined 47 evaluations labeled "empowerment evaluation" published from 1994 through June 2005. They found wide variation among practitioners in adherence to EE principles and weak emphasis on the attainment of empowered outcomes for program beneficiaries. Brad Cousins (2005) wrote a "critical friend" review of EE entitled: "Will the Real Empowerment Evaluator Please Stand Up." Cousins reviewed a number of EE case examples and concluded that any particular EE approach will depend on "which combination of principles are most important, given the needs of the local context and the impetus for the empowerment evaluation in the first place" (p. 201). The 10 empowerment principles are as follows: (1) improvement, (2) community ownership, (3) inclusion, (4) democratic participation, (5) social justice, (6) community knowledge, (7) evidence-based strategies, (8) capacity building, (9) organizational learning, and (10) accountability.

In 2009, Stewart Donaldson as the director of the Claremont University Summer Evaluation Institute organized and moderated a debate about the value of EE between David M. Fetterman, Michael Scriven, and me (Donaldson, Patton, Fetterman, & Scriven, 2010). In the debate, I asked David Fetterman, which of the 10 EE principles he had identified were actually essential. Fetterman responded to my question by explaining that there are "high, medium and low levels of each of these principles . . . You don't always have to do the highest. It depends on your circumstances and situations, but [the principles] give you a gauge and a guide of how to do these things." He continued, "I'm not a purist, not by a long shot. I get along here. You do what you need to do. If you have an idea, theory, a voice, match it up with every day's activities." (Donaldson, Patton, Fetterman, & Scriven, 2010, p. 49) At the time, I didn't consider that a very clear or satisfactory

Table 1. Eight Essential Developmental Evaluation (DE) Principles.

Essential DE principles	What to look for to assess the *degree of manifest sensitivity and sensibility* in DE practice, from design to use of findings	Examples of contextual evidence of the essential DE element being incorporated in practice
1. **Developmental principle**: illuminate, inform, and support what is being developed, by identifying the nature and patterns of *development* (innovation, adaptation, and systems change) and the implications and consequences of those patterns	Something (the innovation) is being *developed*. The evaluation tracks what is being developed and the implications of what emerges. The evaluation itself is developed (emergent design) as the innovation develops	The evaluation's purpose, supporting *development* and adaptation of the innovation, is explicit and that focus is maintained throughout. The evaluation design's emergence and adaptations are documented and their implications discussed
2. **Evaluation rigor principle**: Ask probing evaluation questions; think and engage evaluatively; question assumptions; apply evaluation logic; use appropriate methods; and stay empirically grounded— that is, rigorously gather, interpret, and report data	DE is e*mpirically driven*, and *evaluative thinking* undergirds all aspects of the engagement	Data are gathered, reported, and interpreted about the implications of what is being developed; DE findings and feedback inform next steps in the adaptive process
3. **Utilization-focused principle**: Focus on intended use by intended users from beginning to end, facilitating the evaluation process to ensure utility and actual use	Intended use by intended users focuses the evaluation	Social innovators and their supporters are the primary intended users of DE and clearly identified as such. The explicit purpose of the evaluation is to support the development and adaptation of the innovation (vs. improvement, accountability, or summative judgment)
4. **Innovation niche principle**: Elucidate how the change processes and results being evaluated involve innovation and adaptation, the niche of developmental evaluation	A commitment to innovate is explicit and authentic: a fresh and effective response to an intractable social challenge or problem or to an emergent one	DE has helped the social innovation develop and adapt within the context where the innovation is occurring
5. **Complexity perspective principle**: Understand and interpret development through the lens of complexity and conduct the evaluation accordingly. This means using complexity premises and dynamics to make sense of the problems being addressed; to guide innovation, adaptation, and systems change strategies; to interpret what is developed; to adapt the evaluation design as needed; and to analyze emergent findings	The characteristics of the *complex dynamic system* in which innovation and evaluation are occurring are described. The complexity characteristics of the innovation being developed and evaluated are also described. The DE design, process, and outcomes reflect these complexity characteristics	The nature and degree of uncertainty, turbulence, nonlinear interactions, and dynamical patterns are highlighted. DE is explicitly aligned with the complexity of the innovation. Sensitivity to and implications of emergence, adaptation, and context are manifest

(continued)

Table 1. (continued)

Essential DE principles	What to look for to assess the *degree of manifest sensitivity and sensibility* in DE practice, from design to use of findings	Examples of contextual evidence of the essential DE element being incorporated in practice
6. **Systems thinking principle**: Think systemically throughout, being attentive to interrelationships, perspectives, boundaries, and other key aspects of the social system and context within which the innovation is being developed and the evaluation is being conducted	Attention to interrelationships, perspectives, and boundaries undergirds and informs both the innovation processes and the developmental evaluation	The design, data collected, findings presented, and use of findings demonstrate systems understandings and systems thinking. Contextual sensitivity is explicit and evident throughout the evaluation
7. **Cocreation principle**: Develop the innovation and evaluation together—interwoven, interdependent, iterative, and cocreated—such that the developmental evaluation becomes part of the change process	The developmental evaluator is close enough to the action to build a mutually trusting relationship with the social innovators. The collaborative process is active, reactive, interactive, and adaptive	In the process of collaboration, adaptations and developments are *cocreated*. DE becomes part of the intervention (cocreation). How this occurs and with what implications and consequences are discussed
8. **Timely feedback principle**: Time feedback to inform ongoing adaptation as needs, findings, and insights emerge, rather than only at predetermined times (e.g., quarterly or at midterm and the end of project)	Feedback of findings is timely and ongoing (not just delivered at predetermined times, like quarterly, or midterm and the end of project)	Evidence is reported about how the DE feedback was engaged, useful, and used in close conjunction with real-time decision-making and adaptations

Note. Adapted from Patton (2016).

answer. The latest book on EE (Fetterman, Kaftarian, & Wandersman, 2014) clarifies that there are also *zero levels* of adherence to, implementation of, or attention to the principles. *Zero*. In essence, there are no critical, essential, or core EE principles. It's a pick-and-choose menu. Which ones and how many you engage to merit calling the effort an EE is unspecified on purpose. Part of the EE process is to decide which EE principles to engage and at what level to engage with whatever principles are selected for engagement. In reviewing the new book, I cited as evidence of the pick-and-chose approach Laura Leviton's chapter entitled: "Foundation Strategy Drives the Choice of Empowerment Evaluation Principles."

> Leviton is a former president of the American Evaluation Association and also a Thought Leader in evaluation generally and philanthropic evaluation in particular. In essence, she portrays a foundation-EE relationship that can involve some *hooking-up* now and then around certain mutually attractive principles but ongoing and full commitment is not on the table. Dating, not marriage. Not even co-habiting. Occasional mutual companionship. Always good to have the boundaries of relationships clarified. (Patton, 2015b, p. 16)

Other case examples in the book include some, but not all, of the principles, while several do attend to all, but with varying degrees of emphasis. In responding to this observation, Fetterman, Wandersman, and Kaftarian (2015) explained:

> The essence of EE is a systematic way of thinking, not a single principle, concept, or method. Empowerment evaluation, first and foremost, helps people evaluate their own programs and initiatives. It is the use of evaluation concepts, techniques, and findings to foster improvement and self-determination. It is an evaluation approach that aims to increase the likelihood that programs will achieve results by increasing the capacity of program stakeholders, to plan, implement, and evaluate their own programs.
>
> We presented the theories, concepts, principles, and steps guiding EE. However, his (Patton's) focus was almost exclusively on the principles.
>
> It is the gestalt or whole package that makes it work. Empowerment evaluation theory, concepts, principles, and steps are used to guide practice. Patton's critique is off-target because it focuses on individual parts or principles, failing to recognize that empowerment evaluation is more than the sum of its parts (including "essential" parts; pp. 10–11)

Ah, but that last parenthetical clarification gets to the heart of the matter. *Essential parts. Essential:* "a thing that is absolutely necessary" (Online dictionary, 2015). What is *essential* is never stated. Quite the contrary, a menu of options is offered. Pick-and-choose. The supposed "gestalt or whole package that makes it work" is ultimately ephemeral because the essence is absent. My sense is that in an effort to be inclusive, responsive, and flexible, EE theorists and advocates have created a metaphorical fruit salad of many possible ingredients, none of which is essential, but as long as some of the ingredients are fruit, it can be called a fruit salad. That said, Fetterman (2005) has stated that incorporation of more empowerment principles is better than fewer.

> As a general rule, the quality [of an empowerment evaluation] increases as the number of principles are applied, because they are synergistic. Ideally each of the principles should be enforced at some level. However, specific principles will be more dominant than others in each empowerment evaluation. The principles that dominate will be related to the local context and purpose of evaluation. Not all principles will be adopted equally at any given time or for any given project. (p. 9)

It sounds to me like assessing the degree of manifest sensitivity to principles would work for EE too.

Elaboration of a Sensitizing Concept Approach to the Challenge of Evaluating DE Integrity

Operationalizing a concept involves translating it into concrete measures. This constitutes a well-established, scholarly approach to empirical inquiry. However, concepts like innovation, complexity, emergence, and adaptation are best treated as sensitizing concepts in the tradition of qualitative inquiry, not as operational concepts in the tradition of quantitative research (Patton, 2015a). Since the distinction between sensitizing versus operational concepts is critical to the issue of fidelity and integrity in evaluation approaches, it may be useful to explicate this distinction and its implications for dealing with the essential principles of DE. (This reprises my previous discussion of process use as a sensitizing concept; Patton, 2007.)

Three problems plague operationalization. First, "underdetermination" is the problem of determining "if testable propositions fully operationalize a theory" (Williams, 2004, p. 769). Examples include concepts such as homelessness, self-sufficiency, resilience, and alienation that have variable meanings according to the social context. For example, what "homeless" means varies historically and sociologically. A second problem is that objective scholarly definitions may not capture the subjective definition of those who experience something. Poverty offers an example: What one person considers poverty, another may view as a pretty decent life. The Northwest Area Foundation, which has as its mission "poverty alleviation," has struggled trying to operationalize poverty for outcomes evaluation; moreover, they found that many quite poor people in states such as Iowa and

Montana, who fit every official definition of being in poverty, do not even see themselves as poor, much less "in poverty." Third is the problem of disagreement among social scientists about how to define and operationalize key concepts. Sustainability, for example, can be defined as continuation of a healthy system or the capacity of a system to adapt (Gunderson & Holling, 2002, pp. 27–29; Patton, 2011, p. 199). The second and third problems are related in that one researcher may use a local and context-specific definition to solve the second problem, but this context-specific definition is likely to be different from and conflict with the definition used by other researchers inquiring in other contexts. One way to address problems of operationalization is to treat complexity and innovation as sensitizing concepts and abandon the search for a standardized and universal operational definition. This means that any specific DE would generate a definition that fits the specific context for and purpose of the evaluation.

Sociologist Herbert Blumer (1954) is credited with originating the idea of "sensitizing concept" to orient fieldwork. Sensitizing concepts include notions like victim, stress, stigma, and learning organization that can provide some initial direction to a study as one inquires into how the concept is given meaning in a particular place or set of circumstances (Schwandt, 2001). The observer moves between the sensitizing concept and the real world of social experience, giving shape and substance to the concept and elaborating the conceptual framework with varied manifestations of the concept. Such an approach recognizes that although the specific manifestations of social phenomena vary by time, space, and circumstance, the sensitizing concept is a container for capturing, holding, and examining these manifestations to better understand patterns and implications.

Evaluators commonly use sensitizing concepts to inform their understanding of a situation. Consider the notion of context. Any particular evaluation is designed within some context, and we are admonished to take context into account, be sensitive to context, and watch out for changes in context. But what is context? Systems thinkers posit that system boundaries are inherently arbitrary, so defining what is within the immediate scope of an evaluation versus what is within its surrounding context is inevitably arbitrary, but the distinction is still useful. Indeed, being intentional about deciding what is in the immediate realm of action of an evaluation and what is in the enveloping context can be an illuminating exercise—and stakeholders might well differ in their perspectives. In that sense, the idea of *context* is a sensitizing concept.

High-Inference Versus Low-Inference Variables and Concepts

Another way to think about and understand principles as sensitizing concepts is to treat them as "high-inference concepts." The distinction between high-inference and low-inference variables originated in studies of teacher effectiveness research in higher education (Rosenshine & Furst, 1971).[2]

> High inference teacher characteristics are global, abstract such as "explains clearly" or has good rapport, while low-inference characteristics are specific, concrete teaching behaviors, such as "signals the transition from one topic to the next," and "addresses individual students by name," that can be recorded with very little inference or judgment on the part of a classroom observer. (Murray, 2007, pp. 146–147).

The thrust of the research on teacher effectiveness has been to emphasize low-inference variables that require minimum interpretation as opposed to high-inference variables that require considerable judgment on the part of the observer (Cruickshank & Kennedy, 1986). In contrast, principles-focused practice is necessarily high inference. Addressing degree of manifest sensitivity is a high-inference approach to assessing fidelity and integrity when assessing claims that a particular evaluation approach is being followed.

Table 2. Ten Threats to Evaluation Approach Fidelity and/or Manifest Sensitivity.

Threat	Nature of the threat	Symptoms and problems	Response options
1. Adultery (aka *free love*)	Dabbling with multiple approach partners: start with DE, throw in some theory-driven, a dash of empowerment evaluation, add formative and summative to offer familiarity, heavy infusion of accountability . . .	• Incoherence • Conflicting directions • Surface implementation of each • Confusion • Dissipating resources • Attention deficit, unfocused	✓ Honor the holistic integrity of whatever evaluation approach is chosen ✓ Make an informed, judicious commitment ✓ When using and integrating multiple approaches, which can be appropriate, do so thoughtfully with careful documentation and explicit justification
2. Abstinence	Refusing to even consider using an overall evaluation approach or model	• Focus only on methods • Ethical, utilization, and political issues reduced to technical options and concerns	✓ Ground methods decisions in the larger context of the evaluation's purpose, context, and intended uses
3. Lack of experience (virginity)	Lack the confidence and knowledge to try a new approach for the first time	• Fear • Performance anxiety • Analysis paralysis (can't decide what to do)	✓ Team with a trusted and trustworthy more experienced practitioner ✓ Find a mentor ✓ Start small with a very manageable assignment
4. Impotence	Lack the power, influence, and/or position to advocate for and implement a new approach to evaluation	• Frustration • Boredom (same old, same old all the time) • Rigidity: Forcing new evaluation opportunities and situations into old, comfortable, well-trodden approaches • Stress	✓ Understand the source of the problem. Look for windows of opportunity to demonstrate the value of an alternative approach ✓ Offer exposure to the vast panorama of options and approaches available in evaluation ✓ Don't rush the preparatory work for beginning an evaluation. Take time to understand a particular context and respond sensitively to the needs of stakeholders in that context

(continued)

Table 2. (continued)

Threat	Nature of the threat	Symptoms and problems	Response options
5. Complexus interruptus	Funder of DE or those who commissioned the evaluation want to change course half-way through the evaluation from being complexity- and innovation-driven to being bottom line accountability oriented; or the evaluator is told to stop DE and start doing summative in mid-stream	• Surprised: Didn't see this coming • Anger: Changing the rules in the middle of the game • Concern: How do you recreate the baseline needed for the shift in purpose?	✓ Be vigilant and repetitive about explaining and reaffirming the niche, purpose, and appropriateness of DE to funders and other key stakeholders. ✓ Be especially alert to changes in personnel that will require immediate relationship-building to stay the course
6. Divorce	The evaluation approach isn't working. Advocates and adversaries want to go their separate ways. DE becomes a lightning rod for other issues	• Conflict • Blaming • Scapegoating • Disassociation	✓ Divorce may be better than continuing a bad, conflict-laden, unproductive relationship ✓ Extract lessons and move on
7. Evaluation transmitted disease	Skepticism, even cynicism: "The evaluation approach doesn't matter. In the end it's all pretense and going through the motions. Nobody cares. Just comply with what's required."	• Low energy • Evaluation feels like drudgery—because it is • Lack of cooperation	✓ Begin by acknowledging you have a problem. The problem isn't the way of the world. The problem is you ✓ Get selective. Find an evaluation you can believe in
8. Poor performance	Lack knowledge, competence, commitment, and sensibility to do a good job in appropriately implementing DE.	• DE chosen because it has cache, makes you look cutting edge, generates some buzz • What is done doesn't match core principles of the approach	✓ Commit to a job well done, an approach well-implemented ✓ Identify resources to up your game ✓ Be professional. Take responsibility for adhering to the evaluation standards and principles
9. Poor boundary management	Letting the evaluation expand to include much more than available resources can support. Role confusion around relationships: What are appropriate roles and boundaries for a developmental evaluator?	• Stress • Conflict • Feeling overworked and underappreciated	✓ Realize that this is a common issue, not just about you. ✓ Negotiate boundaries ✓ Better to do a good job on a smaller number of things than a poor job on a lot of things.

(continued)

Table 2. (continued)

Threat	Nature of the threat	Symptoms and problems	Response options
10. Mania	Overselling the approach; overzealous; overapplication. Applying the approach beyond its niche and focused purpose.	• People turn away when they see you coming so they don't have to listen to your inane exhortations about the wonders of this new approach	✓ Better to undersell and overdeliver than to oversell and underperform. ✓ Form relationships with some skeptics and cynics (see Threat 7)—always in large supply in evaluation. ✓ Chill

Note. DE = Developmental Evaluation.

Rigid and Nonrigid Designators

Philosophers of language have devoted considerable attention to different uses of language for different purposes. One critical distinction is between rigid and nonrigid designators. A rigid designator is highly specific and context free, the equivalent of an operational definition or rule. A nonrigid designator depends upon context for meaning such that the interpretation of a term must take into account the situation and the purpose intended by the person speaking. Nonrigid designators apply to the "messy social-psychological world of pragmatics, analyzing the wealth of meaning that must be gleaned not from the words alone but from the context in which the words are produced, including, importantly, the speaker's intentions in uttering them, which furthermore take the speaker outside of his own mind and into the mind of his audience" (Goldstein, 2015, p. 50). Rigid versus nonrigid designators and absolute versus pragmatic (contextual) definitions and meanings take us into the territory of strict constitutional constructionism (focusing on original intent) versus contextual adaptation of interpretation to changing times and situations. A major division on the Supreme Court of United States concerns different interpretations of what *fidelity to the constitution* means. But I digress. Principles as sensitizing concepts are essentially nonrigid designators in the philosophy of language.

Threats to Fidelity, Integrity, and Manifest Sensitivity

Pondering fidelity to evaluation approaches and reflecting on what constitutes integrity in calling an evaluation "developmental" have led me to consider what threats to fidelity and integrity may emerge. I have identified 10 threats: adultery, abstinence, virginity, impotence, *complexus interruptus*, divorce, evaluation transmitted disease, poor performance, poor boundary management, and mania. Table 2 presents the threats and identifies common symptoms. These are serious threats, potentially hiding around every corner. Be afraid. Be very afraid. But also be ready. Table 2 provides strategies for countering the threats.

Conclusion

> A rose is a rose is a rose is a rose.
>
> —Gertrude Stein, Sacred Emily (1913; p. 3)

And a DE is a DE is a DE. Would that it were so. But, actually, it depends. It depends on the extent to which all eight of the essential DE principles have been explicitly and effectively

addressed. That's the point of this article. Labeling an evaluation ''developmental'' or ''utilization-focused'' doesn't make it so. An assessment of manifest sensitivity to a model's essential principles is necessary to judge the integrity of the approach. In closing, let me illustrate why this matters using colonoscopies as an example and cautionary tale.

A colonoscopy is a colonoscopy is a colonoscopy. Or is it? Are there variations? Does it matter how the process is done? A colonoscopy is an examination of the colon with a flexible scope, called an endoscope, to find and cut out any polyps that might cause colon cancer. A study of 12 highly experienced board-certified gastroenterologists in private practice found that some were 10 times better than others at finding adenomas, the polyps that can turn into cancer. One factor distinguishing the more effective from less effective colonoscopies was the amount of time the physician spent examining the colon (which involves an effort evaluation). Those who slowed down and took more time found more polyps. Some completed the procedure in less than 5 min, and others spent 20 min or more. Insurers pay doctors the same no matter how much time they spend. But the stakes are high for patients. More than four million Americans a year have colonoscopies, hoping to protect themselves from colon cancer. The cancer, which kills about 55,000 Americans a year, is the second-leading cause of cancer death in the United States (Kolata, 2006a, 2006b).

So is it true that a colonoscopy is a colonoscopy is a colonoscopy? No more so than that a DE is a DE is a DE. Fidelity matters, whether to prescribed medical practices or evaluation sensitizing concepts. *Caveat emptor.*

Declaration of Conflicting Interests

Funding

Notes

1. Evaluation practitioners involved in developing and/or reviewing the essential DE elements: Mark Cabaj, Nathaniel Foote, Jamie Gamble, Mathias Kjaer, Chi Yan Lam, Kate McKegg, Nora Murphy, Donna Podems, Hallie Preskill, James Radner, Ricardo Ramirez, Rolf Sartorius, Lyn Shulha, Nan Wehipeihana, and Ricardo Wilson-Grau. Different views naturally emerged so the final list of essential elements presented here is my own perspective informed by their feedback.
2. I am indebted to Mel Mark who introduced me to this literature.

References

Blumer, H. (1954). What is wrong with social theory? *American Sociological Review*, *19*, 3–10.

Chilisa, B. (2012). *Indigenous research methodologies*. Thousand Oaks, CA: Sage.

Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, *32*, 199–226.

Cousins, J. B. (2005). Will the real empowerment evaluation please stand up? In D. M. Fetterman & A. Wandersman (Eds.), *Empowerment evaluation principles in practice* (pp. 183–208). New York, NY: Guilford Press.

Cousins, J. B., & Chouinard, J. A. (2012). *Participatory evaluation up close: An integration of research-based knowledge*. Charlotte, NC: Information Age Publishing.

Cousins, J. B., Whitmore, E., & Shulha, L. (2014). Arguments for a common set of principles for collaborative inquiry in evaluation. *American Journal of Evaluation*, *34*, 7–22.

Cruickshank, D. R., & Kennedy, J. J. (1986). Teacher clarity. *Teaching and Teacher Education*, *2*, 43–67.

Daigneault, P. M., & Jacob, S. (2009). Toward accurate measurement of participation: Rethinking the conceptualization and operationalization of participatory evaluation. *American Journal of Evaluation*, *30*, 330–348.

Dickson, R., & Saunders, M. (2014). Developmental evaluation: Lessons for evaluative practice from the SEARCH program. *Evaluation*, *20*, 176–194.

Donaldson, S. I., Patton, M. Q., Fetterman, D., & Scriven, M. (2010). The 2009 Claremont Debates: The promise and pitfalls of utilization-focused and empowerment evaluation. *Journal of Multidisciplinary Evaluation*, *6*, 15–57.

Fetterman, D. M. (2005). A window into the heart and soul of empowerment evaluation. In D. Fetterman & A. Wandersman (Eds.), *Empowerment evaluation principles in practice* (pp. 1–26). New York, NY: Guilford Press.

Fetterman, D. M., Shakeh, J., Kaftarian, S. J., & Wandersman, A. H. (2014). *Empowerment evaluation: Knowledge and tools for self-assessment, evaluation capacity building, and accountability*. Thousand Oaks, CA: Sage.

Fetterman, D. M., Wandersman, A., & Kaftarian, S. (2015). Empowerment evaluation is a systematic way of thinking: A response to Michael Patton. *Evaluation and Program Planning*, *52*, 10–14.

FSG. (2014). Next generation evaluation: Embracing complexity, connectivity, and change. *Stanford Social Innovation Review*. Retrieved from www.ssir.org/nextgenevaluation

Goldstein, R. N. (2015). What philosophers really know. *The New York Review of Books, LXII*, 48–50.

Gunderson, L. H., & Holling, C. S. (Eds.). (2002). *Panarchy: Understanding transformations inhuman and natural system*. Washington, DC: Island Press.

Jacob, S., & Desautels, G. (2014). Assessing the quality of Aboriginal program evaluations. *Canadian Journal of Program Evaluation*, *29*, 62–86.

Kolata, G. (2006a, July 18). Race to the swift? Not necessarily. *New York Times*, Health and Fitness Section, p. 1.

Kolata, G. (2006b, December 14). Study questions colonoscopy effectiveness. *New York Times*, p. A36.

Lam, C. Y., & Shulha, L. M. (2014). Insights on using developmental evaluation for innovating: A case study on the cocreation of an innovative program. *American Journal of Evaluation*, *36*, 358–374.

MacCoy, D. J. (2014). Appreciative inquiry and evaluation – Getting to what works. *Canadian Journal of Program Evaluation*, *29*, 104–127.

Miller, R. L., & Campbell, R. (2006). Taking stock of empowerment evaluation: An empirical review. *American Journal of Evaluation*, *27*, 296–319.

Murray, H. G. (2007). Low inference teaching behaviors and college effectiveness. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 145–183). Dordrecht, The Netherlands: Springer.

Online dictionary. (2015). Retrieved from https://www.google.com/search?q=definition+of+essential&rlz=1C1CHFX_enUS591US591&oq=definition+of+essential+&aqs=chrome.69i57j0l5.9552j0j7&sourceid=chrome&es_sm=0&ie=UTF-8

Patton, M. Q. (2007). Process use as a usefulism. *New Directions for Evaluation*, *116*, 99–112.

Patton, M. Q. (2008). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York, NY: Guilford Press.

Patton, M. Q. (2012). *Essentials of utilization-focused evaluation*. Thousand Oaks, CA: Sage.

Patton, M. Q. (2015a). *Qualitative research and evaluation methods* (4th ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (2015b). Review of empowerment evaluation. *Evaluation and Program Planning*, *52*, 15–18.

Patton, M. Q. (2016). The developmental evaluation mindset: Eight guiding principles. In M. Q. Patton, K. McKegg, & N. Wehipeihana (Eds.), *Developmental evaluation exemplars: Principles in practice* (pp. 289–312). New Yok, NY: Guilford.

Preskill, H., & Beer, T. (2012). Evaluating social innovation. *FSG Webinar*. Retrieved from http://www.fsg. org/publications/evaluating-social-innovation

Preskill, H., & Catsambas, T. T. (2006). *Reframing evaluation through appreciative inquiry*. Thousand Oaks, CA: Sage.

Rosenshine, B., & Furst, N. (1971). Research on teacher performance criteria. In B. Smith (ed.), *Research in teacher education* (pp. 37–72). Englewood Cliffs, NJ: Prentice Hall.

Schwandt, T. A. (2001). *Dictionary of qualitative inquiry* (2nd Rev. ed.). Thousand Oaks, CA: Sage.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (AERA Monograph series on Curriculum Evaluation, *1*, pp. 39–83). Chicago, IL: Rand McNally.

Williams, B. (2005). Systems and systems thinking. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 405–412). Thousand Oaks, CA: Sage.

Williams, B. (2008). Systemic inquiry. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (Vol. *2*, pp. 854–859). Thousand Oaks, CA: Sage.

Williams, B. (2014). *A systems practitioner's journey*. AEA365 blog. Retrieved November 16 from https://us-mg204.mail.yahoo.com/neo/launch?.partner=sbc&.rand=3sva69d24c9fe#

Williams, B., & Hummelbrunner, R. (2011). *Systems concepts in action: A practitioner's toolkit*. Stanford, CA: Stanford University Press.

Williams, B., & Iman, I. (2007). *Systems concepts in evaluation: An expert anthology* (American Evaluation Association Monograph No. 6). Point Reyes, CA: EdgePress.

Williams, B., & van 't Hof, S. (2014). *Wicked solutions: A systems approach to complex problems*. Wellington, New Zealand: Bob Williams. Retrieved from www.bobwilliams.co.nz/wicked.pdf

Williams, M. (2004). ''Operationism/Operationalism.'' In M. S. Lewis-Beck, A. Bryman, & T. Futing Liao (Eds.), *The Sage Encyclopedia of Social Science Research Methods* (pp. 768–778). Thousand Oaks, CA: Sage.