

INCOMPLETE SUCCESSES

Michael Quinn Patton
Utilization-Focused Evaluation
Saint Paul, Minnesota, USA

█ This special issue celebrating the 25th anniversary of the *Canadian Journal of Program Evaluation* shares stories of “how to take advantage of experience with evaluation that cannot be described as successful.” These are what might be called *incomplete successes*. Whenever I hear the phrase “incomplete success” I am reminded that United States President Jimmy Carter used that exact phrase to describe the attempt to rescue Americans held hostage in Iran in 1980. Here is what happened in that now classic exemplar of an incomplete success of gargantuan proportions, one that helps put all other uses of that phrase in perspective.

Operation Eagle Claw was an American military operation ordered by President Carter on April 24, 1980, in an attempt to end the Iran hostage crisis by rescuing the 52 Americans held captive at the U.S. Embassy in Tehran. Eight helicopters with special forces were sent in. Two could not navigate through a very fine sand cloud (a haboob), which resulted in one crash landing and the other returning to the aircraft carrier USS *Nimitz*. Six helicopters reached the initial rendezvous point, but one had damaged its hydraulic systems and could not operate effectively. Spare parts were on one of the helicopters that had aborted. Given these emerging problems, the commanders on the scene requested to abort the mission, which Carter approved. But as the U.S. force prepared to leave Iran, one of the helicopters crashed into a C-130 Hercules transport aircraft containing fuel and a group of servicemen. The resulting fire destroyed the two aircraft involved and resulted in the remaining helicopters being left behind and the deaths of eight American servicemen. No hostages were rescued and none were released until after the election of Ronald Reagan in January 1981. In the interim, Iran dispersed the hostages, making any subsequent rescue attempt impossible. The public debacle damaged American prestige worldwide and is believed by many, including Carter himself, to have played a major role in his defeat in the 1980 presidential election. But that all happened later. On April

29, 1980, when Carter held a press conference to discuss the aborted and disastrous rescue attempt, he called it “an incomplete success.”

When President Barack Obama authorized the incursion into Pakistan that led to the killing of Osama bin Laden in Pakistan on May 2, 2011, Carter’s “incomplete success” was very much on his mind and was often used as a basis of comparison in press coverage. Those involved in planning and executing the Pakistan incursion by U.S. Special Forces reported that they had learned important lessons from the failed Iran hostage rescue about planning and executing such a mission. Retired Chief of Naval Operations Admiral James L. Holloway III led the official investigation in 1980 into the causes of the failure of the operation on behalf of the Joint Chiefs of Staff. The Holloway Report primarily cited “deficiencies in mission planning, command and control, and inter-service operability, and provided a catalyst to reorganize the Department of Defense” (Operation Eagle Claw, 2011).

CRITERIA FOR JUDGING SUCCESS

Obviously, none of the evaluation cases in this special issue reach the level of “incomplete success” exemplified by President Jimmy Carter’s understated judgement. But that example does call our attention to the centrality of criteria in determining what constitutes success. It’s all about criteria. Judgements of merit and worth, success or failure, depend on criteria. The criteria for success (or lack thereof) used by the evaluators who have shared their stories here vary. Let’s look at the various criteria the different authors have emphasized. Then I will comment on some of the cross-cutting patterns that strike me as significant and insightful.

Four of the articles look to the standards of the professional evaluation associations for criteria. Both the article by Françoise Jabot and Murielle Bauchet and the article by Denis Paillard use the French Evaluation Society’s principles as the source for criteria: pluralism, independence (distance), competence, respect for individuals, transparency, responsibility, and opportunity (accountability). Donna Mertens references the Joint Committee’s Standards for Evaluation, the American Evaluation Association Guiding Principles, and the Canadian Evaluation Society’s Competencies for Canadian Evaluation Practice. She comments: “In all of these documents, the evaluator’s relationship with stakeholders and the use of evaluations are important themes. Evaluators who attempt to shape their practice to

align with these competencies, standards, and guidelines find many challenges along the road.” Indran Naidoo opens by noting that the very name of the South African Monitoring Evaluation Association (SAMEA) means that an evaluation must serve multiple purposes, at a minimum both monitoring and evaluation. Other authors focus on criteria that their experiences have taught them to value.

John Owen notes that standards developed by professional societies provide criteria for judging the success of an evaluation, but he finds the scope of those standards “often difficult to implement in practice. For example, how does an evaluator engaged in a complex evaluation keep track of the almost 30 program evaluation standards?” He offers his own areas for judging an evaluation: (a) evaluation plan, including the evaluation design; (b) implementation of the design; (c) dissemination of findings; and (d) evaluation management. But these are topic areas for judging success and not actually criteria for success, although the criteria are implicit in the categories and illustrated in his example. Owen also observes that “evaluation success is a relative concept. In other words, what might be regarded as success for one involved party might not be regarded as such by another.” The criterion he emphasizes to cut through this relativity is methodological rigour that produces findings credible to multiple stakeholders with varying interests. “We had to be sure that the evaluation design was robust enough to support any findings that might not be viewed favourably by these stakeholders.”

Jean-Paul Fortin, Lise Lamothe, and Marie-Pierre Gagnon open their article by stating: “Evaluation use is central to the standards for practice.” Using health evaluations for their focus on evaluation use, they observe that evaluations in Québec have not generally attained high levels of use, and they offer an analysis of why and suggestions for enhancing use.

Linda Lee focuses on the particular challenges of using a participatory approach cross-culturally in coordinating a multi-country evaluation over a period of four years. While authentic and meaningful participation to produce credible evidence are central criteria for judging the processes of participatory evaluation, she judges the ultimate success of the evaluation by its utility. Although there were notable shorter-term uses to influence educational policy in a number of countries, she concludes that

[u]ltimately, the success of this evaluation can be questioned in light of the fact that very little has changed

for Roma children and youth in the region, despite the evidence provided by the evaluation which was used in the European Court of Human Rights and widely disseminated.

Cultural issues also take center stage in the article by Pablo Rodríguez-Bilella and Rafael Monterde-Díaz, which draws lessons from a case study of the evaluation of a rural development program in Argentina, with broader perspectives on the state of evaluation throughout Latin America. A central criterion for them is that an evaluation should move beyond the generation of descriptions to make evaluative judgements about the activities and programs described. To meet this criterion, they conclude, the “culture of evaluation” in the region must be strengthened.

For François Dumaine, the function of evaluation is to determine the extent to which the logic of a program has unfolded as expected. Therefore, his prerequisite criterion for a successful evaluation is to have a program (evaluand) that is adequately defined.

Burt Perrin identifies “good communication” as “one of the key factors contributing to the success of evaluation.”

From my own experience, a key factor to the success of an evaluation is a no-surprises approach, with ongoing communication between the evaluation manager and consultant, both informal and formal. This is a key tenet of a utilization-focused approach (e.g., Patton, 2008). Such contact can help in clarifying expectations, making sure that things are going in the right direction, and agreeing upon changes to the work plan or methods while it is still possible to do so.

Perrin connects his analysis of effective ongoing communications with lessons about the importance of how well the evaluation process is managed.

Penny Hawkins also focuses on how well the evaluation process is managed, more specifically the quality of management of external, independent evaluations, which highlights the criterion of the quality of the relationship between evaluators and their clients, with attention to both the legal contract and shared understandings that go beyond the legal specifications. Thus, she attends to both formal and informal elements of evaluator-client relationships.

With this brief overview of the authors' criteria for judging evaluation success, let me turn to some cross-cutting patterns and themes.

INTENDED USE BY INTENDED USERS: IDENTIFYING AND ATTENDING TO PRIMARY STAKEHOLDERS

All of the articles discuss and illustrate issues that arise in working with stakeholders. Different stakeholders represent different and often conflicting or competing interests. Power differentials and varying expectations can affect everything from evaluation focus to implementation and use. Stakeholders can change or disappear (see the Mertens article). Poor communications between evaluators and stakeholders is a surefire formula for problems (see the Perrin article). Stakeholders become engaged in evaluation with diverse capacities, commitments, openness, and knowledge. Incomplete successes inevitably involve—at some level and to some extent—difficult and challenging relationships with stakeholders, whether those who commission and fund evaluations, those who administer and manage programs, those who administer and manage evaluations, program staff, political advocates of programs, and/or those who are intended to use findings, which may be any or all of the preceding, as well as others. These articles provide an excellent panoramic view of the love-hate relationship evaluators have with stakeholders: can't work with them and can't work without them. Indeed, some of the articles blame incomplete successes on stakeholder inadequacies rather than any evaluator weaknesses, an approach sometimes known as "blame the victim." Many of the recommendations from these experienced evaluators call for more sophisticated and effective ways of working with key stakeholders: engaging them sooner, engaging them with more clarity, building their capacity, managing relationships skillfully, and dealing with problems in a timely way before they become crises. All wise advice.

The importance of focusing on stakeholders, and more particularly, *intended use by intended users* (Patton, 2012), applies to this article. To whom should I direct my reflections on these cases? Beyond the general readership of CJPE, is there a more specific group of primary intended users I might focus on? I've decided there is, and that decision affects how I identify and format my observations and reflections.

I have been much impressed with the annual student case competition of the Canadian Evaluation Society. Cases are an excellent way

of learning how to deal with the complexities and nuances of real-world evaluation practice, which is one of the strengths of this special issue that presents and discusses real-world cases. I co-edited a volume of *New Directions for Evaluation* on “Using the Case Method to Teach Evaluation” (Patrizi & Patton, 2005). I think that this special issue would be a valuable teaching resource, whether with students in training or among a group of evaluation professionals who might engage in reflective practice together using these cases as a starting point to share their own experiences. Reflective practice, it is worth noting, is one of the competency domains of *Competencies for Canadian Evaluation Practice* (Canadian Evaluation Society [CES], 2010). Therefore, what follow are questions that can be considered when using these cases for teaching, learning, and reflective practice.

CROSS-CASE TEACHING AND REFLECTIVE PRACTICE QUESTIONS BASED ON *COMPETENCIES FOR CANADIAN EVALUATION PRACTICE*

1. People Skills

Much evaluation training focuses on learning research methods, which are certainly important, but the fifth and final of the competencies for Canadian evaluation practice is *Interpersonal Practice*. Given my earlier observation about the preponderance of stakeholder relationship issues in the cases, let's start there. Interpersonal practice competencies focus on “people skills, such as communication, negotiation, conflict resolution, collaboration, and diversity.” What specific teaching skills are manifest, as being either present or absent, in the cases? A good place to begin is with Burt Perrin's article which emphasizes that “there is increasing recognition of the importance of ‘soft skills’ such as communications and listening.” Penny Hawkins' article concludes that “the key to a successful or unsuccessful evaluation is often the quality of the relationship between evaluators and their clients.” She opens her article by observing that the “common thread” among the evaluation challenges she analyzes is that “they are all, in the end, about human behaviour and relationships.” That's the general insight. But the devil, as always, is in the details, nowhere more so than in interpersonal relationships.

So here's the challenge: In each article identify at least one specific interpersonal skill that affected the degree of success of that evaluation. For extra credit, assess your own people skills on the specific dimensions you've identified from the articles. For more in-depth

guidance on assessing evaluator skills and competencies as part of determining an evaluator's readiness to conduct a specific evaluation, see Patton (2012, step 2).

2. Context Considerations

Françoise Jabot and Murielle Bauchet describe an evaluation experience in health promotion in which the evaluator must adapt her practice to the situation and conclude that “the context is fundamental here.” The geo-cultural contexts in these articles vary globally, including Québec, Argentina, Australia, east central and southeastern Europe, France, South Africa, Canada, and the United States. The settings in which the evaluations are conducted include national governments, educational institutions, nongovernmental programs, a philanthropic foundation initiative, and multinational agencies. But these differentiations are too broad to provide much insight into the importance of context. The theme of the annual meeting of the American Evaluation Association in 2009 under the presidency of Debra Rog was “Context.” All agreed that context matters greatly, but a wide variety of definitions of context offered varying interpretations of both how to assess context and how to take it into account in designing and implementing an evaluation.

So here's the challenge: In each case presented in this issue, identify how context affected the degree of success of the evaluation. Be specific. Use this analysis to generate and deepen your own sense of how context affects evaluation. This is a core element in the Situational Practice category of the *Competencies for Canadian Evaluation Practice*, which includes attending to the unique “contextual circumstances in which evaluation skills are being applied.” For more in-depth guidance about situation analysis, see Patton (2012, step one).

3. Management Practice

The competencies in this category focus on “the process of managing a project/evaluation, such as budgeting, coordinating resources and supervising” (CES, 2010, p. 2). Naidoo's article offers profound and important insights about the challenges of establishing and managing a national monitoring and evaluation system. Hawkins focuses her article on managing evaluation contracts. Perrin laments that there is very little attention in the literature to the competencies required by evaluation managers about how best to manage the evaluation process. Lee's article describes the challenges of manag-

ing a diverse international evaluation team. All of the articles raise and illustrate issues related to managing evaluation. Identify at least one central evaluation management issue in each article that directly affected the degree of success of the evaluation.

4. Technical Practice

At the root of *incomplete evaluation successes* we may find people problems (stakeholder and interpersonal interaction issues), contextual challenges, and evaluation management inadequacies. All of these elements ultimately affect the quality of the design, the credibility of data collected, the appropriateness of the analysis, and the overall technical rigour of the evaluation. Technical practice competencies focus on these specialized aspects of evaluation, such as design, data collection, analysis, interpretation, and reporting (CES, 2010). So it's time to turn to methodological analysis. Owen, in his article, is especially focused on how technical quality affects credibility of findings. Naidoo shows how lack of technical competence among evaluators can undermine an entire M&E system. Lee discusses the challenges of variations in assessment tools and reporting standards. Indeed, since these are evaluation cases, they all involve technical and methodological issues.

The challenge for this category is twofold. First, identify at least one important technical or methodological issue that affected the degree of success of each evaluation presented in this special issue. That's the easy part. The second challenge is to describe and analyze how that technical or methodological issue relates to and is a result of one or more of the following: stakeholder issues, contextual factors, and evaluation management practicalities. This involves moving from constructing a simple list of success factors to understanding the complex dynamic interactions among those factors, dimensions, and competencies. (For more on applying a complexity framing to evaluation, see Patton 2011, 2012.)

ADVANCED AND SPECIALIZED REFLECTIVE PRACTICE QUESTIONS

The preceding questions are based on and engage the groundbreaking *Competencies for Canadian Evaluation Practice*. One of the tests of the competencies framework is the extent to which they provide a framework for informing professional development. Using the com-

petency categories to examine, compare, contrast, and understand these cases of incomplete evaluation success offers a great opportunity to apply the competencies in support of reflective practice and see how well they work. I found that having them explicitly in mind as I read the cases generated important insights about the challenges in these cases as well as in evaluation practice more generally. What I propose to do now is add some additional specific questions that arose for me from the cases.

5. Control Issues

Evaluators do not control evaluations. Many would like to. Many aspire to. Many actually try to. But if one takes the standard of utility seriously, we as evaluators are not the primary intended users of evaluations, so we don't control the ultimate outcome of high quality evaluations: intended use by intended users. We can attempt to create the conditions for use, get users primed for use, support their capacity to be knowledgeable users, and facilitate use—but, ultimately, we don't control use. We are only enablers.

And evaluation use is just the proverbial tip of the iceberg of what we don't control. Denis Paillard examines an evaluation “that combines a number of the problems outlined in a textbook example: confusion between ‘control’ and ‘evaluation’.” This is but one explicit and insightful manifestation of control issues. As all the cases make clear, evaluators often need approval from others before they can finalize the design, measurement instruments, analysis procedures, reporting times, report format, dissemination of findings, and budgetary expenditures. The frustration and anguish that screams from some of these evaluators, and is clearly present but buried beneath the surface in others, is this fact of evaluation life: *We are not in control.*

I invite you to test this premise and my assertion of its importance and implications. In each case, identify a central control issue. How does the evaluator's lack of control affect the evaluation's incomplete success? What recommendations flow from these experienced, insightful, and dare I say courageous evaluation colleagues, about dealing with lack of control? Examine closely. How many practice recommendations take the form of trying to exercise more control? Keep in mind that dealing with lack of control and uncertainty are core constructs for handling complex situations and endeavours (Patton, 2011), complexity being a cross-cutting theme of these cases, in my judgement. (Note to students: If you are a control freak, this

is a good time to reexamine evaluation as a career choice. Consider instead becoming an air traffic controller.)

6. Conflicting Purposes

I have long argued, usually in vain, that evaluations should be focused on and have a clear, distinct, explicit, and agreed-on purpose. Just as no single evaluation can serve everyone's interests and answer everyone's questions (thus the centrality of identifying major stakeholders and primary intended users), no single evaluation can well serve multiple and conflicting purposes. Choices have to be made and priorities established. But instead, evaluators, those who commission and fund evaluations, and intended users persist in expecting evaluations to serve multiple and competing purposes equally well. Those unrealistic expectations, and the failure to set priorities, are often the place where evaluations start down the long and perilous road to incomplete success.

Consider this example. Naidoo opens the deeply engaging and insightful South African case by stating that M&E must be undertaken in a manner that addresses "multiple M&E objectives, and recognize[s] that it also contributes to broader objectives of supporting transformation and improving governance." The case then goes on to describe the capacity limitations of doing even one thing well, much less achieving multiple objectives like accountability, program improvement, and adaptive decision-making. Transformation and improved governance? Supporting either of these ideals is challenging. Supporting them both within a single M&E system? Count me as skeptical.

Yet the rhetoric of multiple purposes is rampant. I see evaluation proposal after evaluation proposal that purports to ensure accountability, improve programs (formative evaluation), render overall judgements of merit and worth (summative evaluation), generate generalizable lessons, and—oh yes—contribute to developmental adaptations. So here's the assignment: Examine each case for centrality of evaluation purpose. To what extent do you find that an important element of incomplete success is trying to do too much by serving multiple purposes and competing stakeholders' needs with limited resources? Which evaluation cases, in contrast, manifest singularity of purpose? How did that singularity of purpose constitute a success factor or a barrier to success, or proved irrelevant? And in your own evaluation practice, how skilled are you at presenting purpose

options, the implications of those options, and facilitating setting priorities?

7. Power and Politics

I fear that we have become cavalier in repeating the mantra, as John Owen does early in his case, that all evaluation is political. Politics and power issues go hand in glove, and these cases provide plenty of evidence of both power inequities in relationships and political considerations in all aspects of evaluation, from determining priority questions to deciding what to disseminate. To help us avoid treating power and politics as mere contextual truisms and to assist us to in avoiding a cavalier acknowledgement of the political nature of evaluation, these cases offer an opportunity to dig deep and get real. Each case manifests power relationship dynamics, either explicitly or implicitly. Each case presents political considerations as a factor in its incomplete success.

As I have urged throughout this review, let's again get specific. For each case, identify at least one power inequity and analyze its influence on the evaluation's process and use. Do likewise for one critical political consideration in each case. Complete this analysis by examining how power inequities and political considerations interact to affect some core element of the evaluation's incomplete success. In some cases you'll need to read between the lines and go beneath the surface. It is often the nature of power politics that the power and politics are hidden from view. But they always become visible at some point in an evaluation, at least if you're paying attention. Practice paying attention in these instructive teaching cases.

8. Evaluation Culture and Evaluative Thinking

Pablo Rodríguez-Bilella and Rafael Monterde-Díaz conclude that the "culture of evaluation" in South America must be strengthened. The *Competencies for Canadian Evaluation Practice* include the application of evaluative thinking. Linda Lee observes:

As we pay more attention to cultures when we cross borders, as evaluators we also need to recognize different cultures within our own, seemingly more familiar, context. Culture itself is a dense and multilayered phenomenon, the effects of which add another dimension of

unpredictability to evaluation and, ultimately, the influence or success of the evaluation process and its results.

As attention to the importance of evaluation culture and evaluative thinking has increased, we face the danger that these phrases will become vacuous through sheer repetition and lip service. Here, again, the cases provide an opportunity to get real, to test our understanding of what is meant by evaluation culture and evaluative thinking, and how these two critical concepts are interconnected and cross-fertilizing. Every case presents either explicitly or implicitly some assumptions about what constitutes evaluation culture and evaluative thinking, and how these sensitizing constructs provide a foundation for evaluation success, and, contrariwise, how their absence increases the likelihood of evaluation failure. Systematically examine the cases to capture what is meant by these constructs when discussed explicitly, and examine basic assumptions in the other cases to bring to light what is implied about these constructs in the cases where they are not discussed explicitly. Use this analysis to sharpen your own understanding and definition of these critical ideas.

WHY SUCCESS ALWAYS STARTS WITH FAILURE

I want to close by expressing appreciation to CJPE editor Brad Cousins and the special issue's guest editor Marie Gervais for their courageous vision in conceptualizing this 25th anniversary issue, and to the courageous and insightful authors who responded to the challenge of remembering, sharing and offering insights into taking advantage of "an experience evaluation that cannot be described as successful." We give lip service to learning from failure and encourage those with whom we work to do so. We are, therefore, called on to walk the talk by both acknowledging our incomplete successes and learning from them. This special issue walks the talk, a fitting focus for CJPE's quarter-century anniversary.

Tim Harford (2011) has written a book, *Adapt: Why Success Always Starts with Failure*, that examines why this special issue is so important for the future of the evaluation profession. In the book, he draws on a broad range of historical and recent examples to illustrate the importance of making effective use of failures. He argues for the importance of systematically learning from trial and error, expecting to engage in evolution as failure teaches how to adapt, and building success through learning and adaptation. In my view, the predominant lesson in this issue's cases is the importance of adapta-

tion, as an evaluation unfolds from early negotiations about focus through methods decisions and data collection through to reporting and follow-up for use. Learning from incomplete successes of the kind offered in this special issue is essential to evaluation's future long-term success as a profession.

REFERENCES

- Canadian Evaluation Society. (2010). *Competencies for Canadian evaluation practice*. Retrieved from http://www.evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice.pdf
- Harford, T. (2011). *Adapt: Why success always starts with failure*. New York, NY: Farrar, Straus and Giroux.
- Operation Eagle Claw. (2011). *Operation Eagle Claw*. Retrieved from http://en.wikipedia.org/wiki/Operation_Eagle_Claw
- Patrizi, P., & Patton, M. (Eds.). (2005). Using the case method to teach evaluation. *New Directions for Evaluation*, No. 105.
- 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.

Michael Quinn Patton is an independent evaluation consultant with 40 years experience. He is former president of the American Evaluation Association and author of *Essentials of Utilization-Focused Evaluation* (2012) and *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use* (2011). He teaches regularly in The Evaluators Institute and the World Bank's International Program in Development Evaluation Training (IP-DET).