

REFLECTIONS OVER 25 YEARS: EVALUATION THEN, NOW, AND INTO THE FUTURE

Ross Conner

Professor Emeritus, School of Social Ecology
University of California Irvine

When approached to write a commentary for this volume, I was very excited to be involved because I value failure. Let me explain. I am a firm believer that we can learn more from our failures than from our successes. I say “*can* learn more” instead of “*do* learn more” because, too often, we miss this valuable opportunity, focusing instead on the frustration, defeat, misunderstandings, recriminations, and justifications that are part of the experience. But failure, or “mistakes” more generally, lie in the realm of knowledge, so if we are alert, we get closer to better knowledge.¹ That was the promise and potential I foresaw in the contributions to this special 25th anniversary issue and why I quickly said “yes!” without knowing what the contents would be or who the authors would be.

Let me give you an example of why I value failure and mistakes. During my college years at Johns Hopkins University in Baltimore, Maryland, I volunteered as a tutor for students from Baltimore’s very poor inner-city neighbourhood. One of the elementary school students I tutored, Billy, needed help with math, particularly fractions. I would instruct him on how fractions worked, and then watch as he tackled a problem. Clearly, Billy was working hard and he always came up with an answer—but it was the wrong answer. I tried instructing again, he tried calculating again, but again the result was the wrong answer. So I switched the roles: I let Billy teach me his method of working with fractions. His method was very systematic and thoughtful but it was not the “right” method. Once I understood and could reproduce his method (something Billy was quite pleased about, since he had successfully taught me), I was able to see why he did what he did and what caused his errors. Once we were anchored in his “mistakes,” we were able to move together to the correct way to handle fractions. It was his mistakes, once we focused on them, that led Billy—and me—to better knowledge.

Corresponding author: Ross Conner, School of Social Ecology, University of California, Irvine, CA, USA 92651; <rfconner@uci.edu>

The charge to the authors in this issue was “to share some evaluation experiences that have evolved unexpectedly, creating turbulences in the evaluation process and threatening the very success of the evaluation.” The 15 authors of the 11 articles in this issue have met this challenge admirably. Some of the articles read like exciting short stories, where you cannot wait to turn the page to learn more. Others are more restrained in their tone but equally rich with experiences and lessons. Together, they provide a way for us to see what challenges skilled evaluators are facing at this point in evaluation’s development, how these evaluators reacted to those challenges, how they frame their analyses about the “failure,” and what the lessons for greater knowledge are as we move into the future.

Although not selected for this reason, the articles in this special 25th anniversary issue of the *Canadian Journal of Program Evaluation* provide a good picture of where evaluation is now, and they also demonstrate how much evaluation has changed in the past 25 years, not only in Canada but also elsewhere around the world. In reaching this conclusion, I revisited the first CJPE issue from Spring 1986 to get a snapshot of evaluation at that time. I also drew upon my recent experiences as president of the International Organisation for Cooperation in Evaluation (IOCE), the worldwide “United Nations-like” group of national and regional evaluation organizations, groups, and networks, and my international evaluation experiences since then. During my presidential term and in the experiences after, I have had the honour and pleasure of meeting with evaluators around the world, learning more about the significant successes and the varied challenges that the evaluation field faces, and working with many different evaluators and their supporters to advance evaluation into the future.

THEN AND NOW

Comparing the contents of the first CJPE issue 25 years ago and this issue, I was immediately struck by several differences. First, the language of evaluation has become multilingual. In the first issue of CJPE, all but 1 of the 11 articles were prepared in English.² For the 10 English-language articles, there was a nod to Canada’s English-French context, with a short abstract in French but no English abstract. The 1 article in French, following the same pattern, had a short English abstract but no abstract in French. In the current issue, in contrast, 7 articles are in English and 4 are in French. One of the English articles, however, was originally in Spanish (*Evaluación*,

*Valoración, Negociación: Reflexiones en camino hacia una cultura de la evaluación*³); the authors decided to have it translated into English to aid the CJPE readers.

Second, evaluation has become multicultural, reaching many parts of the globe. The authors of the articles 25 years ago were, not surprisingly, predominantly Canadians.⁴ Nine of the 11 authors/coauthors were Canadian; the other 2 were American. In this current issue, Canadians still predominate but just barely: 5 authors or coauthors are from Canada, 4 are from France, and 1 each are from Argentina, Australia, New Zealand, South Africa, Spain, and the United States. This issue, therefore, has perspectives from five continents; the first issue had only one, from North America.

Third, evaluation involves a balance of men and women, something that was less true 25 years ago. In the first edition, 9 out of 10 authors were men; only 1 was a woman. In this current issue, there are 8 males and 7 females. As I experienced the evaluation world while IOCE president and since, it is clear that women are active and involved in important positions at all levels in our field. For example, the current president of the IOCE is a woman, as are several other leaders of evaluation associations around the world (e.g., AfrEA, the African Evaluation Association).

The important fourth difference relates to content. I will comment on this aspect below, as part of my discussion of the themes that cut across the articles in this issue.

NOW AND THEN

As I undertook my review of the articles in this 25th anniversary issue, three topics emerged and then took on new and more complex aspects as I reread each article's examples and cases and reflected on the contents of the first issue 25 years earlier. These topics were communication, standards/competencies, and methodology.

Communication

This important theme appears in every article in this issue but rarely arises in the CJPE issues from 25 years ago. In the first CJPE issue, there was some discussion of generic resistance to evaluation; in this 25th-year issue, in contrast, the focus is much more personal. The discussion focuses on the dynamics and quality of the relationship

between the evaluator and the other stakeholders who are directly or indirectly related to some or all stages in the evaluation process. In all cases presented here, it is the problematic quality of the communication that is the issue. This is not surprising in a set of papers focusing on failures and, in fact, is very telling. When communication is not good, it presents serious threats to the success of an evaluation.

Many of the evaluator-stakeholder relationships described in these papers were fraught with problems. For example, in the case described by Rodríguez-Bilella and Monterde-Díaz, the Supervisor of Program Evaluation, after reading a draft of the final evaluation report, characterized the evaluators as “very angry” about the results from their observations and asked them to delete “opinions or inferences” and to take “comparisons out of the report.” The Supervisor tried to explain the reason for his concerns, telling the evaluators that “there is not a culture of monitoring and evaluation as a way to improve performance. In general, especially by the authorities, monitoring and evaluation are considered as a ‘trial’ to punish, that is why I am asking for these changes.” The evaluators revised their report three more times, but the Supervisor continued to have problems with it, calling the evaluators too evaluative in their statements and criticizing them for making “potentially hurtful” statements. In this case, the communication problems were multifaceted, as is typically the case. The Supervisor did not understand what evaluation is, what evaluators do, and how the process of evaluation works. The evaluators also did not understand the Supervisor’s concerns about “hurtful” evaluation findings and his reluctance to undertake any evaluation activities beyond simple descriptions.

In the Fortin, Lamothe, and Gagnon case, there were also misunderstandings about evaluation’s processes and outcomes. The learnings and lessons from the evaluations of several health IT projects were lost on, or perhaps ignored by, the planners of similar future projects, even though the evaluations were strong, well-planned, and well-executed. They discuss “human factors” as one important cause of the problems.

Other writers describe communication problems of a different sort. Both Hawkins and Mertens describe cases where communication problems resulted from major personnel shifts during the evaluation process, causing new stakeholders to be in critical positions. In Hawkins’ case, for example, different contract managers were in place during the two most critical steps in the five-step evaluation

management process: commissioning and reporting. The lack of good communication between the two different contract managers and with the evaluator threatened the success of the evaluation when the evaluators delivered their report. Hawkins explains why, during the commissioning stage, the “spirit” of the discussions between evaluator and client is as important, and perhaps more important, than the “letter” of agreement, commonly expressed in Terms of Reference, Memoranda of Understanding, Evaluation Plan, Process Descriptions, and so on. It is a good “social contract,” a positive working relationship, that is important, Hawkins argues, particularly when unexpected and challenging developments occur during the course of the evaluation. In Mertens’ case, “none of the people who were on the list [of stakeholders] at the beginning of the evaluation continued to hold those same positions by the end of the project.” In Hawkins’ terms, the original social contracts were severed; Mertens and her team rethought the possible project users and beneficiaries and then designed, implemented, and reported their evaluation for them.

Jabot and Bauchet also identify communication issues as a major need for successful evaluation, and they lament that it is not one of the seven principles that the French Evaluation Society (Société française de l'évaluation, SFÉ) sets out: “le partage des informations tout au long du processus est une composante majeure pour l'appropriation des résultats, mais la communication ne figure pas parmi les principes.”⁵ They conclude, with an exclamation, that communication is an indispensable ingredient: “La communication entre les différents acteurs est un ingrédient indispensable au succès d'une évaluation!”⁶

In her article, Lee describes working in east central and southeastern Europe on projects focused on educational programs for the Roma youth populations in four countries. Because she did not speak the languages of the area, she recognized from the outset that communication would be a challenge. She did not anticipate, however, that miscommunication would exist on several levels beyond language. Lee says, “I have begun to understand that I am frequently working in contexts that are foreign to me although, on the surface, they appear familiar.” She describes “one of the most critical moments for the evaluation (and indeed the whole project)” that demonstrate the power, for better or for worse, of communication. “Not only was hope of a successful evaluation evaporating before my eyes,” Lee reports, “the entire project appeared to be in jeopardy.” The discussion involved all the important stakeholders: project directors, teachers, and

evaluators/researchers, and, through the raw and frank comments of one evaluator/researcher, resulted in some stakeholders exchanging heated comments and others in tears (you can read more details in Lee's article).

Owen highlights the importance of communication as well. In the case he describes, about a project to improve the quality of teaching in schools, there were three important stakeholder groups: the program developer, the regional directors of the program, and the policy directors from the central office of the educational system. The evaluators had "frank and direct" communication with the last group of stakeholders, in contrast to their communication with the first two stakeholder groups. This turned out to be important when the findings showed that the program was fatally flawed in design and hence not achieving its objectives. Although the reaction to the evaluation results by the program developer was "apoplectic" and by the regional directors was "unhappy," the central office policy directors were more accepting of the negative results, albeit still "squeamish" about how their bosses would react. Given the strong and clear negative evaluation results and the recommendation for the termination of the program, this would be a challenging situation for any evaluator, no matter what the quality of communication among the stakeholders. In fact, Owen concludes that "it is likely that all three major stakeholders felt that this had been an unsuccessful review because the findings had shown that the program that they had developed, sponsored, or championed had been shown to be flawed." Nonetheless, Owen and his team's good communications with at least one of the important stakeholder groups provided an anchor from which to move forward.

Dumaine highlights another benefit of good communication: it facilitates a rich and deep discussion between the evaluator and the key stakeholders at the points in the evaluation process where the critical decisions are made during the negotiation about the evaluation's focus, approach, and emphasis. Although Dumaine focuses on the benefits of evaluability assessment rather than specifically on communication, he recognizes that good communication is critical because it is only through a meaningful dialogue among the key stakeholders that the important realities related to and the general approaches of the evaluation process emerge and therefore make the negotiations successful. Citing Rossi et al., Dumaine notes that "la qualité première de l'étude d'évaluabilité est de reconnaître que les paramètres d'une évaluation, incluant la définition de son objet, doivent faire l'objet d'une négociation entre les parties prenantes."⁷

Naidoo makes a similar point about the inevitability of negotiation when planning and conducting an evaluation and the need for good communication to make it successful, productive, and useful. In explaining the lack of use and uptake of evaluation results in the South African Department of Land Affairs, Naidoo believes the evaluators in the department's M&E Directorate were at fault because communication was not good with other stakeholders in the department and they "failed to properly negotiate around the organizational dynamics, and thus [were] not able to generate sufficient levels of organizational learning."

Finally, Perrin focused directly on the critical importance of good communication, even including this phrase in his article's title. In his four examples, Perrin demonstrates how "communication-gone-wrong," using his term, leads to a variety of challenges for the evaluator. Communication-gone-right, in contrast, "can help facilitate a collaborative and collegial, rather than an adversarial, approach between the external evaluator and internal project manager that invariably leads to synergies, cooperation, maximizing institutional knowledge, and greater relevance and use."

Evaluation Standards/Evaluator Competencies

All but one of the current authors included some focus on evaluation standards, in stark contrast to 25 years ago when standards did not exist and the need for them did not concern evaluators. The standards that were mentioned most were those from the Canadian Evaluation Society, the French Evaluation Society (SFE), and the American Evaluation Association, although others were also referenced. It is striking not only that so many writers mentioned standards in their papers, but also that a set of standards served as the major framework for analyzing evaluation failure in three papers. Jabot and Bauchet used the seven principles in the French standards to deconstruct their community health promotion evaluation, and Paillard used the same set of principles to analyze the failure of the evaluation of two regional development programs. The seven principles in the French standards are Plurality (of perspectives), Detachment (impartiality), Competence, Respect for People, Transparency, Opportunity, and Responsibility. Jabot and Bauchet concluded that the most important characteristic explaining their evaluation failure was missing from the seven SFE principles: communication. Paillard, in analyzing his "évaluation ratée"⁸ against the SFE principles, concludes, "aucun des principes figurant dans cette chartre n'a été respecté."⁹

Owen also uses standards to assess the evaluation he and his team conducted, although in a novel way. After noting that standards are often difficult to implement in practice, he groups them into four stages related to the evaluation process: planning, implementation, dissemination, and overall management. Owen believes the evaluator's conduct during the first stage, evaluation planning, is the most important in fostering a successful evaluation, with good, honest communication at the core. Applying this idea to his own evaluation case, he concludes, "In retrospect this might be one component of our evaluation approach in the ISP case that was not successful."

Other authors also include mention of standards in their papers, sometimes in direct analysis and other times in passing references. Nonetheless, the frequent inclusion of standards indicates the extent to which, in the past 25 years, standards and the associated issues of evaluator certification and credentialing have become central to the evaluation field. During my term as IOCE president, I was frequently asked about standards, certification, and credentialing, particularly by evaluators in developing regions of the world. I typically referred people to the Canadian Evaluation Society website for more information because CES was and still is the leader in collecting, synthesizing, and recommending next steps about professional competencies for evaluators.

In my view, standards may be useful to evaluators, but I am dubious about how relevant they are to the many other key stakeholders in the evaluation process. What some authors in this issue wanted, implicitly, was a set of standards that those commissioning or funding evaluations should and would follow. Standards can be useful to guide evaluators, and the certification/credentialing of evaluators can raise evaluator competencies and may be helpful to evaluation commissioners and funders to select well-trained evaluators. Unfortunately, when it is time for decisions and actions in an evaluation project, the current standards and the certificates do not provide the anchors that are needed. The focus reverts to the human factors—the communication among evaluators and stakeholders, the main topic among all this issue's writers.

Methodology

The absence of much methodological discussion struck me in this set of articles. In contrast, Vol. 1, No. 1 had several very methodologically focused articles, with some surprising statements that are as true

then as they are today.¹⁰ In the current papers, the main sources of failure that authors identify are not, for instance, a flawed design or a poorly constructed survey.

This shift away from a central focus on methodology indicates two things. First, because of the development of multiple methodologies over the past few decades, we now have a much larger set of options to answer a wider variety of evaluation questions. So we are better equipped than we were in the past. Second, the initial and central focus for an evaluator beginning an evaluation needs to be on the evaluation questions, not on the ways to answer them, which is what methodology focuses on. It is good to see that all of these authors focus on better communication between evaluators and the key stakeholders as ways to identify the correct evaluation questions, to adjust those questions and the associated methodologies as needed during the process of doing the evaluation, and to share answers to the questions, particularly when those answers are not what the stakeholders expected or hoped.

We seem to be moving away from having the same methodological discussions and disagreements among ourselves, and instead are paying more attention to the larger world of evaluation stakeholders, who move forward with us or sometimes without us (e.g., in the current worldwide focus on impact evaluation in development, the economists are controlling the discussion, with evaluators largely on the outside looking in).

INTO THE FUTURE

What will evaluation look like 25 years from now? That is a question that I am not foolish enough to try to answer, but I will venture a few ideas about trends that could or should happen in the next few years, based on what the evaluators in this issue have reported and also on my recent experiences with those around the world working in the area of evaluation.

Better communication will continue to be an important issue. As all the authors here make clear, when communication is not good, it presents serious threats to the success of an evaluation. But is the reverse true—that when communication is good, it results in a successful evaluation? I believe it *facilitates* but does not guarantee a successful evaluation. Evaluations are complex endeavours, so there needs to be good communication among all the stakeholders

over a long period, and, as new realities—including new stakeholders—emerge, the evaluator needs to be skilled enough to make wise adjustments and then to keep making them. This is where the trend toward standards, credentials, and certifications will assist.

Professional standards and professional designations will continue to be important trends in evaluation. As we move forward in these areas, we need to look to other similar professions that have moved further along this path. Accountants, for example, undertake some work that is very similar to evaluation and, in some instances, is identical to evaluation. In much of the developed world and some of the developing world as well, accountants have standards and codes that are widely available and widely recognized by all participating stakeholders, something that is not true for evaluators. Accountants can undertake different types of assessments, and clients/stakeholders understand what the parameters and obligations are, on all sides, for each different type. This is an area where evaluation needs to make more progress. Evaluators understand and agree among themselves, more or less, on what constitutes good standards and practices. Other stakeholders in the evaluation process, however, are largely unaware and generally uninterested in these standards and practices; they simply want evaluators to evaluate. Until these other stakeholders recognize and accept the standards and best practices and also accept the obligations on their parts that go with them, the power of having them will not be available to evaluators.

Looking back 25 years at the first issue of CJPE demonstrates the strides forward we have made but also highlights the questions that still linger. The only article in the first issue published both in English and in French was entitled “In Defense of Program Evaluation.” It was written by John Mayne, then president of CES, who is still active in the field and is well known to many of us. He was responding to a series of news articles about the just-released *Final Report* of the Ministerial Task Force on Program Review, essentially a government-wide review of the status of program evaluation across Canada in 1986. Mayne writes that many of the newspaper articles “had disparaging things to say about federal program evaluation: references to the general poor quality of evaluations and ‘useless’ and ‘self-serving’ evaluations abound.” In program evaluation’s defense, Mayne refutes these claims and provides good evidence to the contrary, from the subreports that the *Final Report* was based on.

At this point in evaluation's history, we have probably moved beyond the point where a broadside attack of this sort could be made on evaluation. But we still have work to do. At the end of his comments, Mayne asks a series of questions that we still have not resolved:

Should evaluations cover a given set of issues? Should evaluations only use state of the art methods and always be as definitive as possible? Should a reader be able to pick up an evaluation and expect to find a certain type of information? You know what your [sic] going to get out of a financial statement and sort of what you are going to get out of an audit—operational problems identified. Should there be a parallel with an evaluation report? Can evaluations both serve management and provide a basis for public accountability?

Mayne's last question in particular hangs in the air unresolved, 25 years later. We can and should answer it, or at least make an attempt at answering it, and, depending on the answer, begin to specify much more clearly how it can be done—or how differently focused evaluations, those for management or those for public accountability, are done. We probably will make some mistakes, but that is not a bad thing if, like the authors in this issue, we are open to learning from our mistakes and failures. As I found with Billy, my elementary school tutee in Baltimore years ago, it is failures that lead us forward to knowledge.

NOTES

- 1 By coincidence, as I was completing my final revisions, the *New York Times* featured an article by Paul Tough, "What if the Secret to Success is Failure?" (September 14, 2011). http://www.nytimes.com/2011/09/18/magazine/what-if-the-secret-to-success-is-failure.html?_r=1&ref=magazine
- 2 I had an incomplete electronic copy of the contents to review, possibly with one article missing, so there might have been an additional paper and it might have been in French. Also, the final article was unique in being presented in full text in both English and French, although it likely was originally written in English (based on my personal acquaintance with its author).

- 3 *Evaluation, Assessment, Negotiation: Reflections on the Way Towards a Culture of Assessment.*
- 4 Also of note, several of the authors were in high-level positions in government and in the CES, e.g., the Comptroller General of Canada, several from the Office of the Auditor General of Canada, and the then-current president of the Canadian Evaluation Society, John Mayne.
- 5 Information-sharing throughout the process is a major factor in the ownership of the results, yet communication is not among the charter principles.
- 6 The communication between the various players is an essential ingredient for the success of an evaluation!
- 7 The key to evaluability assessment is the recognition that the evaluation parameters, including the stated objective of the evaluation, have to be negotiated among the stakeholders.
- 8 Botched evaluation.
- 9 None of the principles enumerated in the charter was followed.
- 10 For example, in the article by John Crane, he comments on “the inability of traditional group comparison designs to fulfill the great expectations of them that were held three decades ago.” After 55 years, therefore, we are still arguing about the applicability and utility of control group designs for program evaluation.

Ross Conner is Professor Emeritus, Social Ecology School, University of California, Irvine, USA. His evaluation work has focused on a variety of sectors including health, education, criminal justice, and leadership. He has published nine books, with two more in progress, plus numerous papers/reports on evaluation project results and on politics and evaluation, evaluation utilization in program improvement and policy formation, evaluation design, international and cross-cultural evaluation issues, and evaluation training. He is former president of the American Evaluation Association and former president of the International Organisation for Cooperation in Evaluation.