SUCCESSFUL OR NOT: IT DEPENDS ON YOUR FRAME OF REFERENCE

John M. Owen
Centre for Program Evaluation, The University of Melbourne
Parkville, Australia

Abstract: This article describes an evaluation that was judged to be unsuccessful from the point of view of key program stakeholders. This was due to the fact that the evaluation did not support program advocates who had much to gain from positive evaluation findings. We argue that, although the knowledge needs of stakeholders must be taken into account, the integrity of evaluation practice, consistent with codes of professional behaviour, should guide the conduct of all aspects of what evaluators do.

Résumé :   Cet article décrit une évaluation jugée infructueuse du point de vue des parties prenantes clés du programme. Selon l'auteur, cette situation s'explique par le fait que l'évaluation n'a pas favorisé les défenseurs du programme qui avaient beaucoup à gagner de résultats positifs de l'évaluation. Bien que les besoins d'information des parties prenantes doivent être pris en compte, l'auteur soutient que l'intégrité, conformément aux codes de conduite professionnelle, doit guider la conduite de l'évaluateur dans tous les aspects de sa pratique.

INTRODUCTION

A few years ago I was the leader of an evaluation team that was asked to evaluate an intervention designed to improve the quality of teaching in schools in an educational system. The intervention, which I will call the Improving Schools Program (ISP), was in the prototype stage, in that it was being trialled in approximately 20 schools in one educational region within the system. The ISP was a complex intervention designed to engender school-level commitment to improved teaching and learning. An external evaluation was commissioned six months after the ISP commenced, and was to run for a period of eight months. The commissioning agent was keen to obtain an indication of the program worth, and so would be consistent with

---

Corresponding author: John M. Owen, Principal Fellow, Centre for Program Evaluation, The University of Melbourne, Parkville, Australia, 3010; <j.owen@unimelb.edu.au>

the principles of impact evaluation. This was despite the fact that the program would have not run its course by the end of the evaluation period. This is a challenge that many evaluators face, when policy makers' timelines for decision making cannot wait for an intervention to fully run its course or be fully implemented.

The evaluation is a useful vehicle for the purpose of adhering to the theme of this issue of CJPE—that is, to explore notions associated with the success or otherwise of evaluation practice. The advanced organizer for this discussion is 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. This was the situation for the ISP evaluation, where indicators of program worth provided a challenge given the timing and timeline of the evaluation.

THE STAKEHOLDERS

Most evaluation studies involve multiple stakeholders, parties that have a legitimate interest in an intervention under review. In this case, there were three key stakeholders in the ISP: the program developer, a director of education of a region in the system, and policy developers in central office in the system. A short description of the context and interests of each stakeholder is necessary to understand how the evaluation findings had the potential to affect them.

The program developer (hereafter PD) did not have a traditional background in educational theory, having come from a background in systems analysis. He had latterly developed an interest in education and school change; the ISP was a practical manifestation of his ideas. A charismatic figure, he had persuaded some in the educational bureaucracy that the adoption of the ISP would lead to a significant improvement in schooling. *If widespread adoption followed the trial, there would be significant financial advantages for him over the longer term.*

The regional director (hereafter DR) allowed the trial to take place in his region. With a reputation for getting things done, and for knocking heads together if necessary, he was committed to a broad five-year plan to revolutionize schools in the region. The ISP was one program designed to achieve the goals of the plan. As preparation for the trial of the ISP, there had been extensive consultations involving the PD, meetings with school principals,

and training for regional consultants. *If the ISP were seen to work, there would be kudos for the DR among his peers in the educational system.*

The policy developers (hereafter ES) worked in the central office of the educational system. Their role was to advise senior management about the implementation of policy across regions and schools. The ISP had come to their attention through very early developments of the program in other jurisdictions. The PD had also made contacts with these policy makers to persuade them to implement the ISP across the entire school system. However, it was important that the ISP was consistent with the most recent policy directions. At this time, schools had been bombarded with initiatives and there had been some resistance from principals that they were being asked to do too much. *The ES had to be sure that the ISP would not upset broader policy change imperatives set by the government of the day.*

As Ernie House has said, all evaluation is political (House, 1993). Findings from an evaluation have the potential to reward some stakeholders and disadvantage others. In the ISP, a favourable evaluation would mean good news for all stakeholders. With this in mind, 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.

THE EVALUATION

Given their concern to ensure that the ISP was consistent with broad policy, the ES could be regarded as the most neutral of the stakeholders in terms of program advocacy. This was consistent with the notion that middle-ranking members of an organization need to be sure that changes they advocate will be consistent with systemic goals and direction. This helps us understand why the ES not only commissioned the evaluation, but also saw themselves as the primary audience for the study. Understanding this context allowed the evaluation team to develop a professional rapport with the ES, which led to frank and direct conversations about what was being found as the evaluation progressed.

Consistent with good evaluation practice, the evaluation team developed an evaluation plan that took into account the key concern of the ES, and linked methodologies to the key evaluation issues (Owen, 2006). The issues were

1. *The nature of the ISP*. What are the assumptions, objectives, and strategies of the program; to what extent is it internally coherent?
2. *Consistency with the knowledge base about whole school improvement*. To what extent is the ISP consistent with relevant research on good practice? To what extent is the program plausible in the context of this educational system?
3. *Translation of the program into practice*. How is the program being implemented? To what extent does the program lead to improved teacher practice? What is the relative advantage of the ISP compared to alternative programs used by schools?
4. *The future of the ISP*. Should the program be adopted more widely? If so, to what extent, and how should it be modified?

There is no space or reason to go into detail about the methodologies used in this study. However, the evaluation team subscribed to the notion of pragmatic evaluation in the selection of these methodologies (Datta, 1997). This involves the creative use of complementary mixed-method designs to get as "close to reality" as possible given the exigencies created by timing and resource constraints of the evaluation.

Key aspects consistent with a pragmatic approach in this evaluation were to

- involve all stakeholders in the development of the ISP program logic. As seems to be the case in the majority of programmatic interventions (Bickman 2000), this had not been done by the program developers. We found that none of the stakeholders had ever heard of program logic.[1] A benefit of this process was that the program developer and regional director felt they were being involved in the evaluation.
- use an expert panel to compare the logic, manifest in the assumptions and expected outcomes of the ISP, with the current research on educational change and improvement. In this exercise the evaluation team made no contribution to the discussion, except to structure the session and take notes of what was said by the panel.
- allow regional consultants, principals, and teachers to comment on early implementation of the ISP. A key finding was that inadequate change strategies of the developer as part of the ISP meant that there was a lack of clarity in schools about what the program was meant to implement and how

it was meant to be implemented. There was also an opportunity for these sources to provide opinions about whether the program would produce student outcomes consistent with system-level objectives.

THE FINDINGS

There were two key findings from the analyses. The first was that the program relied on a four-stage logic model. Each stage in itself had a series of processes that led to outcomes for that stage. The operations of the third and fourth stages were conditional on the achievement of the outcomes of the first and second stages. In no school in the trial were the outcomes of the first and second stage achieved satisfactorily. The second finding involved responses from the expert panel. In summary, they could not reconcile the assumptions in the logic with what was known about effective school improvement.

These findings had a quality assurance orientation about them, and fitted in well with the fact that the evaluation was undertaken during the first eight months of program implementation.

It was clear that the ISP was flawed in design and, not surprisingly, schools were having a hard time putting the program into practice (issues 2 and 3 above). On the basis of these findings, the evaluation team reported that the program was unlikely to contribute to system-wide objectives and we decided to recommend that it be discontinued.

SUCCESSFUL EVALUATION OR NOT: STAKEHOLDERS' PERSPECTIVES

It has long been accepted that promoting use of findings is an integral part of good evaluation practice, and studies that I have been associated with have been based on effective dissemination practices (Alkin & Taut, 2003). In this case a report was written for the policy makers (ES), the primary audience for the study. We also held a seminar to which the policy makers invited the program developer (PD) and the regional director (DR), and at which the findings of the report were discussed.

To say the least, these meetings were stressful.

The PD was apoplectic, launching a prepared defence of his program, and in particular attacking the program logic as one basis for our

decision. The fact that we had asked him to validate the logic as a strategy during the implementation of the evaluation design provided a strong counterargument to his claims. In addition to railing against us, he also attacked the policy makers. He had expected this to be a "formative" evaluation, when in fact the policy makers had apparently agreed to an evaluation that took a more critical perspective. *From the perspective of the PD, this was an unsuccessful evaluation.*

The DR and his deputy were likewise unhappy with the findings. They took particular issue with the make-up of the expert panel, and demanded information about problems of implementation in the sample schools. It was clear that the findings were at variance with what the DR wanted from the evaluation, a surprise that had to be communicated to consultants in the regional office, and a hurdle to the ongoing commitment to the program in the region. *From the perspective of the DR, this was an unsuccessful evaluation.*

Finally, what about the two policy makers (ES)? The fact that we had been in constant touch with the ES during the life of the evaluation cushioned the impact of the findings that was felt by the other key program stakeholders. Nevertheless there was considerable squeamishness about how they would present the report to "senior policy makers upstairs," as considerable resources had been used to fund the ISP trial. One can imagine that a positive set of findings would have been more comfortable for these officers. As we know, programs tend to develop a life of their own. An evaluation that recommends the cessation of an intervention presents a range of problems for policy makers.

Such decision-making was, fortunately, not our responsibility, which ended at the conceptual stage in the knowledge-to-use continuum. However, there were several indications that we had changed the reference frame of the ES, consistent with the view that "if information changes (a policy maker's) preferences or understandings.... utilization is a reality. Altering frames of reference is important because, in the long run, the policy-maker's new vision will emerge in different policy priorities" (Knott & Wildavsky, 1980, p. 303). *So the jury is out as to whether this evaluation was also unsuccessful from the point of view of these policy makers.*

SUCCESSFUL OR NOT: THE EVALUATORS' PERSPECTIVE

The discussion above implies that one way to judge the success of an evaluation is to assess the effects of the findings on program

stakeholders. In this case study, 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.

What of the evaluation team? Our view was that we *had* undertaken a successful evaluation even though we had not provided findings that had been wanted by the stakeholders.

While standards developed by professional societies have provided guidance to practicing evaluators and to judge the success of an evaluation, it is my view that they are 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 developed through the auspices of the American Evaluation Association (Sanders, 1995)? One way to manage this complexity is to judge an evaluation on the basis of four key elements (Neale, Owen, & Small, 2003): the evaluation plan, including the evaluation design; implementation of the design; dissemination of findings; and evaluation management.

So, in recognition of the dangers of self-evaluation, we briefly reflect on these elements in regard to the ISP case, as evidence for making the claim that the ISP evaluation was successful.

Evaluation plan/design. Consultation led to a comprehensive plan that included a pragmatic study design that acknowledged time and timing limitations. The design linked key questions to methods that could provide plausible answers to these questions. Past experiences of the evaluation team and access to relevant theory were essential for implementing pragmatic principles.

Implementation. Arrangements were made to collect and analyze data without major hiccups. The ethical aspect of data collection was observed. Attention to timelines and quality control across evaluation team was essential.

Dissemination. The primary audience was kept informed about evaluation. The written report documented the argument for discontinuation of program. There was opportunity for program stakeholders to question conclusions and voice dissatisfaction about the findings. An ability of evaluation to justify findings in the face of pressure to compromise on results was essential.

Evaluation management. The evaluation team was savvy about context, developed understandings about general policy directions, realized the importance of an external frame of reference for making judgements of worth, and involved stakeholders in key evaluation processes without compromising its independence on data collection that led to adverse program findings. Understanding the political nature of evaluation was essential.

I contend that all four elements need to be judged as successful for an evaluation to be successful. For example, an evaluation could be judged as unsuccessful if the findings do not reach the intended audiences ("Dissemination"), even if the design was of the highest quality. This was a problem in the early days of evaluation when dissemination was confined to the end-of-evaluation report.

Some in the evaluation profession perceive an inherent tension between responding to the information needs of clients and adhering at the same time to relevant codes of practice. Many commissioners of evaluation tacitly acknowledge that, when asking for an external review of an intervention for which they are responsible, an evaluator often brings such a code. But this is not always the case; there is ample evidence that many practicing evaluators are not aware that these codes exist.

During the evaluation planning stage of an evaluation (the first element above), an understanding needs to be set regarding the conduct of a given study. It is my view that evaluations have the most chance of success if the evaluator plays a leadership role in this stage. This involves keeping in mind professional criteria for success while searching out the "real" information needs of clients. This might mean that the information needs of some clients cannot be met, while others may not be of consequence. And this also includes opening a dialogue about the possibility that an evaluation will provide findings that are not in the best interests of the stakeholders concerned. In retrospect this might be one component of our evaluation approach in the ISP case that was not successful.

NOTE

1   A spin-off from this evaluation was that the policy branch of the educational system became heavily committed to program logic in other interventions they were responsible for.

REFERENCES

Alkin, M. C., & Taut, S. M. (2003). Unbundling evaluation use. *Studies in Educational Evaluation, 29*, 1–12.

Bickman, L. E. (2000). Summing up program theory. *New Directions for Program Evaluation, 87,* 103–112.

Datta, L. (1997). A pragmatic basis for mixed method designs. *New Directions in Program Evaluation, 74*(Summer), 33–46.

House, E. R. (1993). *Professional evaluation: Social impact and political consequences*. Thousand Oaks, CA: Sage.

Knott, J., & Wildavsky, A. (1980). If dissemination is the solution, what is the problem? *Knowledge: Creation, Dissemination, Utilization, 1*(4), 537–575.

Neale, J., Owen, J., & Small, D. (2003). Encouraging the use of codes of behaviour in evaluation practice. *Evaluation and Program Planning, 26,* 29–36.

Owen, J. (2006). *Program evaluation: Forms and approaches.* Sydney, Australia: Allen and Unwin.

Sanders, J. R. (1995). Standards and principles. *New Directions for Program Evaluation, 66*(Summer), 47–52.

John Owen is interested in evaluation as a form of systematic knowledge production, and the use of this knowledge in effective policy and organizational decision making. He is a Fellow of the Australasian Evaluation Society. Recently he has been engaged in supporting groups in countries that are newly engaged in evaluation practice.