

THROUGH THE LOOKING GLASS:
WHAT HAPPENS WHEN AN
EVALUATOR'S PROGRAM IS EVALUATED
and
DEGREES OF STRANGENESS

Alan G. Ryan and Lynne E. Baillie

█ Evaluators have been exhorted to “make the familiar strange” and not merely to re-create the existing structures and relationships they have been asked to study. Perhaps an even greater challenge is to “make the self strange”: to unpack the personal baggage that we carry into the research process and that influences our interpretations.

From the unfamiliar positions of “client” and “evaluator” within an institution itself grappling with constructions of identity, the authors of the following two articles became “strangers” as they found their perceptions threatened both by the new roles in which they found themselves and by the forces of organizational memory that surrounded the new program.

The client, Alan, is an evaluator of many years. He is also Assistant Dean of Undergraduate Programs for the College of Education, University of Saskatchewan. His responsibilities include administering the undergraduate program implemented four years prior to his appointment and overseeing its evaluation. To the latter end, Alan and an advisory panel selected an evaluator to conduct the relevant research.

The evaluator, Lynne, has evaluated and designed programs in the private sector for several years. After receiving her doctorate under Alan's supervision, she worked as a sessional lecturer and as a supervisor in the college before the revised program was implemented.

The program discussed in these articles is a revised undergraduate program implemented four years ago at the College of Education. Its intention was to offer a sequence of new courses that would encourage student teachers to construct their own understanding of

becoming a teacher. In the process of implementing this sequence, college faculty members were required to develop many new courses, revise or delete all others, and design a variety of new school-based experiences to complement the courses.

The articles that follow were written independently and are presented as such to provide two parallel but quite different perspectives on one evaluation. The authors, standing in unfamiliar roles within familiar territory, found that this new and uncomfortable vantage point not only gave them insights into their organization and their colleagues, but also challenged their own understanding of the evaluation process.

THROUGH THE LOOKING GLASS: WHAT HAPPENS WHEN AN EVALUATOR'S PROGRAM IS EVALUATED

Alan G. Ryan
College of Education
University of Saskatchewan

Abstract: This paper reports on an occasion where the author, an experienced evaluator, became the client of an evaluation. This reversal led the author to reflect on the role of the client in an evaluation. The article records the thoughts and impressions of a client as his program was being evaluated. The discussion section focuses on the roles and obligations of, respectively, the client and the evaluator.

Résumé: Cet article rend compte d'une occasion où l'auteur, un évaluateur expérimenté, se trouvait dans le rôle de client d'une évaluation. Ce revirement a permis à l'auteur de réfléchir sur le rôle de client d'évaluation. L'article présente les pensées du client et les effets sur le client, lors de l'évaluation de son programme. La section de discussion concentre l'attention sur les rôles et les obligations du client et de l'évaluateur respectivement.

An evaluator can be useful to both program planners and implementers, but tensions can arise from the differing interests and purposes of these various parties. (Rossi & Freeman, 1989, p. 113)

Rossi and Freeman (1989), and indeed many others who write on evaluation, remind evaluators of the importance of sensitivity to the needs and expectations of their clients. In such writings, however, the client is often presented as a depersonalized, anonymous figure who requires the evaluator's consideration and attention, not as a person in his or her own right. To avoid such abstraction, Alkin and Associates (1985) encourage evaluators to put themselves in the client's position in order to understand what would constitute a meaningful evaluation report from the point of view of the client. This article goes one step further; it reports on the thoughts and impressions of a client as his program was being evaluated.

The program undergoing evaluation and the persons involved (Lynne, the evaluator, and Alan, the client) were described in the foregoing introduction. This article is in the nature of a commentary on the interaction of client and evaluator from the perspective of an experienced evaluator who, for a change, was himself the client of an evaluation. It is largely based on extracts from a journal that I kept over the course of the evaluation. In the discussion, the roles and obligations of the client and the evaluator, respectively, are considered.

THE EVALUATOR BEGINS WORK

Lynne developed her proposal in November 1995. She worked with documents that I made available to her, and we consulted along the way. Initially, I was somewhat concerned about the impact of the evaluation report. I already knew that the new program had problems, some of them due to the magnitude of the changes the new program introduced and some of them of a systemic nature, mainly stemming from budget cutbacks. I wrote in my journal:

I don't think that I will feel comfortable with giving total control to Lynne. I'm not sure if she is politically sensitive enough as an evaluator to ensure that the report isn't a minefield. But I do have to give her more room, that is for sure. We meet on Friday and I think that we should have this issue out or she will be very uncertain where she stands.

After that meeting, the journal entry reads:

Very reassuring meeting. We discovered that we were on much the same track with respect to how to conceptualize the evaluation. I think she wants to make it broader than I do. For example, she wouldn't mind including the student advising system, whereas I feel that that is not really part of the new program. But we are agreed on having to dig below the surface, e.g., just because students say there is overlap between courses, we need to see what form that overlap takes. We agreed on March as a reasonable aiming point for bringing closure to the project. One of the political dimensions we will have to be aware of is the potential for mischief of whatever we say. I still say "we," I notice, but really more and more I see it as Lynne's study.

Nothing much happened before the Christmas break. The Undergraduate Program Committee (UPC) approved Lynne's proposal, and she began to develop her plan of action for the new year. Initially, I had conceived of the evaluation as a joint effort, where we would be co-evaluators. But my role of co-evaluator was not to be; as events unfolded, the pressures of my new position made it impossible to take an active role in the evaluation, even if Lynne had encouraged that. I notice in my journal that at one point I mentioned helping in aspects of the study such as document analysis but that I was unsure how Lynne had received my offer. In the end, I limited my involvement to conducting two case studies with two students who were in their last term of the new program. These case studies were incorporated as a component of the final report.

THE TWO CASE STUDIES

Although the two case studies were taking shape at the same time as the rest of the evaluation study was moving ahead, I will discuss them completely in this section before returning to the main agenda of this article: my interaction as client with an evaluator. The principal reason for mentioning the case studies at all is that they became an important counterbalance, in terms of both their content and their emotional importance to me, to Lynne's final report.

I approached two students whom I had taught early in their programs. The first, who chose the pseudonym "Kim," was in the secondary education stream. She was a direct entry student, which meant that she had entered the college directly from high school

and would take four years to obtain her B.Ed. degree. The second student, who chose to be called “Kyra,” was in the elementary education stream. She was a post-academic student; that is, she had completed her arts and science courses before she entered the college for the two-year post-academic route to the B.Ed. I wanted an honest reflection of student experiences in the new program; consequently, I chose students who I knew were intelligent and thoughtful, but who were independent thinkers and thus not necessarily supportive of all that happened to them in their time with us.

The methodology was very straightforward. In a series of interviews, I took each student through her experiences in the college, from start to the present. Both students were highly articulate, and so my contributions to the interviews were minimal. I transcribed the tapes and from them constructed the two case studies. There was no thematic analysis; the stories were simply organized chronologically. They were each just under 20 pages long, and so there was considerable description and reflection. I returned the original transcripts and the case studies to each participant and obtained the usual member checking, changes in the text, and signing off. Both students expressed satisfaction at having been involved in the evaluation and were pleased with the way their views were presented in the written documents.

From my perspective, the case studies extended the study in two ways. First, they gave a more longitudinal portrayal than the evaluation study alone could. For example, reading Kim’s study, I was struck by her mentioning that a course she felt was over her head in her first year affected her significantly in her third and fourth years. Reading the case studies leaves the impression that the students had created something organic, evolving, and personal out of their college experiences.

The second way in which the case studies contributed was in terms of Scriven’s (1967) distinction between *payoff* evaluation and *intrinsic* evaluation. Lynne was concerned with intrinsic matters: how the courses are being delivered, how the courses fit together, and how the conceptual objectives of the program are being met. Scriven, ever the consumer advocate, also acknowledges payoff evaluation, that is, evaluation that concentrates on what customers feel about a product and how it meets their needs. As the two students looked back at the totality of their experiences, they were in essence painting a picture of how the program met their needs in becoming teachers. We will return to this point during the discussion portion of this article.

INTERACTIONS DURING THE EVALUATION

Lynne and I met fairly frequently during the months of her study. Here is a typical entry from the journal of this period:

Had coffee with Lynne this morning. Talking about some of her first impressions. Lots of interesting stuff, obviously, but still she is operating on a meta-level and I do hope that I can persuade her to make some very specific recommendations about specific issues. It is good that she is sharing with me a bit of what she is finding. I don't think I am influencing her in any particular direction — come to think of it, I don't really think there is a particular direction I want to influence her in.

Lynne and I have always had an easy and open working relationship. Even when we were involved in Lynne's doctoral work, we met more as equals than as student and mentor. I sometimes wondered how others perceived our roles in the evaluation. I wrote:

One thought to add: I wonder how Lynne and I are perceived by faculty as we discuss things over coffee — am I seen as subverting the process? Or am I becoming paranoid?

DRAWING TOWARD THE CLOSE

As Lynne ended her data collection and analysis phase, my attention as client was drawn back to the study. I began to wonder what my "rights" as a client were and, the obverse, what I could not impose lest I compromise the study. In my journal I wrote:

Soon after I wrote the last entry, she came to me and said that as she was about to start the report, we should now meet on a fairly regular basis so that the report could be designed to meet my (my personal, or my, the College?) needs. I know that Lynne's preference is to write a "lofty" report, by which I mean one that provides overviews and analyses at an organizational or philosophical level, rather than just a bunch of nitty gritty recommendations. I think that that will be valuable, certainly for people like the new dean who won't know the history of the program. But I also know that faculty are

waiting for specific answers, no matter how hard I have been trying to downplay to them how specific and targeted an evaluation can be. So my pitch to Lynne will be to have some of both. And as a client, I think I am entitled to specify the areas where I (and here I am speaking not only on my own behalf but also on the behalf of other interested people who are eagerly awaiting the report, UPC, for example) really need the outsider's perspective so that we have something against which to react. That's the key point, isn't it, the need for something solid against which to react. Even if one doesn't agree with a recommendation, one can at least use it to react against. Without that, really, the report cannot be as powerful.

After the meeting I added this short note:

Had a good meeting with Lynne. I needn't have worried about leading her — she has very clear ideas of what she wants to say. I limited my contributions to suggesting that she put a positive spin on what she says about the program being limited by personnel currently in place. I suggested that she think of it in terms of "opportunities to be addressed when hiring." I also reiterated my need for direct answers to specific questions. I think she will do those anyway, but it was important for me to remind her.

During the period when Lynne was crafting the final report, she and I would informally go over some of the issues that she was starting to target. Looking back on the journal entries, I can sense the beginning of some unease:

Lynne wants to talk about her findings and I must admit that I am interested in learning about them. Perhaps it interferes with the separation of our two roles but on the other hand, seeing it from the point of view of an evaluator meeting with her client, surely it is good strategy to meet with your client before the final report is done so that you can break the news gently. If that is what we do, then I can't see anything wrong. But there is no doubt that I do want to influence her in some respects because I am the one who has to make maximum use of the findings so I want them in a form that helps me the most.

Later

A most interesting chat with Lynne. She is finding lots of problems with the program, both at a conceptual level and at a delivery level. I didn't put any of my values on her findings. Lynne would like me to monitor drafts of the report so that it can be written in a way that is most useful for people who have to act on its findings. She has no doubt that there are systemic problems with the program that are pulling it down; I think so, too, but I don't think that I believe that they are pulling it down as far as she does. I offered to react to a "pilot" issue, i.e., to have Lynne take one small issue and treat it in the way that she feels the report should be, and I could react to that and perhaps that would help her in crafting the total report. But she felt that it would be better if she kept it at a "total" level; she already feels that she has too many little issues and what she needs is to develop an overarching design of the report so that it makes sense in some totality. Whatever, it won't be a dull document!

About this time, wider political issues were starting to intrude. Primarily, the concern of the university community was with the prospect of massive downsizing as a result of federal and provincial government cuts. Much of the speculation around campus was whether certain colleges of the University of Saskatchewan would merge with those of the University of Regina. The two colleges most commonly mentioned were those of Education and Engineering. And the popular saw-off among the public was to say, "Let the University of Regina take Education and the University of Saskatchewan take Engineering." The impact of such a move on the students of the province, and on the whole University of Saskatchewan, would be fairly drastic, never mind the effect on current students and staff of the colleges concerned. But these are interesting times; and although the informed view is that rational decision making would not countenance such an ill-considered move, it is also clear that these are not altogether rational times. The journal reflects my unease:

Just had my meeting with Lynne. It looks like her report will be more far ranging than I anticipated. She wants to examine the whole organization and how it handles change. Wonderful stuff, but as she spoke, what flashed into my mind was the VP Finance's remarks in the pa-

per this morning about whether we need two colleges of education in the province. This is not a good climate into which to drop a bombshell like this [report]. But yet these things need to be said. I explained my concerns to Lynne and I don't know that we reached an agreement, because I didn't really want my concerns to dominate her agenda, but yet I do have to worry about the implications in the wider community of her type of report.

THE FINAL REPORT

Lynne delivered the draft of her report over the weekend, and I was shocked. Here is what I wrote at that darkest time:

Well, Lynne delivered the first draft of the report on Sunday and I read it in one gulp. My first reaction was total despair. My second reaction is pretty much total despair, too. There doesn't seem to be one good word to say about the program in her report. I have nothing on which to build. It could set faction against faction, and give everybody the right to claim victim status. I was really desolate, I must admit.

When I said I have nothing on which to build, I meant that if the purpose of evaluation is to improve, not to prove, I don't know where to start. Lynne talks about the diminished resources available in the College and yet most of her recommendations would require more efforts placed on the few willing shoulders left. Mentoring students and faculty, for example. Great ideas, but certainly the load involved in mentoring students is horrendous. Perhaps mentoring faculty is more feasible, if we had any faculty to mentor. But who is going to mentor whom, and to what end?

I am intrigued with my own reaction. I was trying to think if I would have felt this way if I had been a regular faculty member still. I have to admit I wouldn't. Indeed, I would probably have taken the same perverse pleasure in venting that so many of the interviewees took. It made me realize just how much I am not a regular faculty member any more, I am the person in charge of the program. On one level, I am awed by the amount

of work addressing the report will take, but that doesn't really worry me. I guess what worries me more is that I won't be able to do it, that I won't be able to keep the exercise from spinning out of control and causing the disintegration of something that I have invested a lot of my years in. And that I find scary. And then sometimes I swing the other way totally, and say: "Why not just put the report out there, let it clear the air, and let's see what happens." I think I would be comfortable with that, if I felt that the air-clearing would be constructive and not destructive, and I am not convinced of that.

From an evaluation perspective, I think that the big difference between Lynne and me is that I am always conscious that people, even administrators, have to be left with some good feelings so that they can be positively disposed to build on my reports. I always make sure that something positive gets said; I write a much more political document. Programs have to be salvaged, you don't shut down many programs totally. Life goes on. I don't know if this difference in attitude is a consequence of my greater experience doing evaluations, or whether it is a reflection of our different personalities. That is hard to say, and in the end perhaps it doesn't matter.

What to do? Well, of course I am going to talk with Lynne about it. As a minimum, I want her to clean up quotes so that clearly wrong information isn't left in there. I guess what is troubling me is that I have to reconcile this portrait of a totally dysfunctional program with the feedback from Directors that they feel that our graduates are better prepared now than they have ever been. I know our program has some problems; indeed, I don't think that Lynne uncovered any that I didn't know about. But I know that our program has strengths, too.

Lynne and I met for a lengthy discussion of the report. I explained that I was really worried that the report could threaten the very existence of the program, and she maintained that her findings were so pervasive that she could do nothing else but report them as she found them. She agreed to revisit her report and see if there was anything she could do to ameliorate my concerns while at the same time staying true to her data. She kept her word; notwithstanding, the final report carried the same overall message.

As the days went by, I became more sanguine about the report's impact on the continued existence of the college. Partly, this was a natural reaction to my initial despair — I don't stay negative for long. Second, it became apparent that it would not be possible to release the report until June at the earliest, and by that time the minister's special advisor would probably be winding up his findings. The main reason for the delay was the difficulty in collecting some portions of the statistical data and the slowness of obtaining sign-off permissions for the case studies (the students were writing examinations and were not in the mood to read their studies). Also, I took my holidays for the month of May. Third, I also began to consider how I could use the report for improving the program; I had always appreciated the report's potential for this, of course, but that aspect had been overshadowed by the greater concern of radical political decisions.

DISCUSSION

The object of the exercise was achieved: I now knew what it was like to be a client of a program evaluation and, perhaps more significantly, a client who receives a negative report about his program. I now turn to a brief analysis of the relationship between client and evaluator, in terms of the rights and obligations of each.

The Rights and Obligations of the Client

The client commissions the evaluation. Without that commission, there would be no study, no findings, no report. The client is purchasing a service. As such, the client has the right to expect certain things: that the report will be timely, honest, methodologically sound, and so forth (Braskamp, Brandenburg, & Ory, 1987; Guba, 1975). But does the client also have obligations? Is the client required to keep a distance from the evaluation, or may the client intervene? Meyers (1981) directs the client to ensure evaluators are not harried:

There is a possibility that the evaluators might be coerced or co-opted by the close relationship that is likely to grow between them and the administrator ... The administrator must take extreme care that evaluators are not co-opted or discouraged from pursuing any lines of inquiry. (p. 84)

May the client suppress the report? Ignore it? Change or distort it? Or does the report, once produced, have a right to its existence similar to the way that a baby, once born, is deemed to have that right? As an evaluator, I feel drawn to this latter position. A client need not be forced to call for an evaluation, but once one is called for and delivered, the client does not have the right to put the report to death, and so become privileged with respect to that particular piece of knowledge. As Cronbach and Associates (1980) state, "An open society becomes a closed society when only the officials know what is going on" (p. 4). Watson (1987) writes:

The consultant's report should not disappear or be dismembered because managers do not like the methods or conclusions. The only way to get objective advice is to ask for it and to let it stand when it is received. Reject the advice if necessary, but do not hide it. (p. 83)

As the client in this instance, I still intend, on principle, to hold to the view that the report I was instrumental in commissioning must live. Yet that said, it is a fact that Lynne and I had an extensive meeting after I read the first draft of the report during which I deliberately and consciously tried to get her to see certain issues in the same light that I did, with the purpose of getting her to change some aspects of the report. As a simple example, I, the client, strongly objected to the use of the word "dire" in the description "A program in dire need of focus and direction." In my mind, that translated too neatly into a juicy headline in the local newspaper. Lynne probably never caught the implication of the clichéd adjective in the way that I did. When I pointed it out, she agreed to delete the word and probably wondered what all the fuss was about.

A more serious issue arose out of Lynne's extensive use of quotations from interviews. Although I am a great believer in the power of quotations to represent stakeholders' value positions, I became worried as I noticed my reaction to their onslaught. I respect Lynne's position (a valid one, I believe) that people's perceptions are what need to be addressed, regardless of whether they are factually correct. If a person believes that "they" are out to get him or her, then that perception needs to be addressed. However, some of the quotations contained outright errors; the people being quoted had just got the facts wrong. I wanted these corrected and Lynne didn't want to touch them. To her, they were perceptions and therefore legitimate to include; to me, they created an erroneous impression of the program in a reader's mind. They were irritants that I was going to

have to deal with even though I knew they were false. I felt that the line between “I think that this is wrong with the program” (a perception) and “This is wrong with the program” (a substantive shortcoming) was becoming blurred in the reader’s mind. Consequently, a perception, no matter how incorrect or silly, could take on the status of a substantive shortcoming. In the end, some of the contentious quotations were adjusted and others were not, but this was an instance where I felt that my rights as the client were being compromised. Of course, Lynne probably felt that her rights as evaluator were similarly being compromised.

The Rights and Obligations of the Evaluator

The obligations of the evaluator are well known. They are perhaps most fully enshrined in the formal standards (Joint Committee on Standards for Educational Evaluation, 1994). The evaluator’s rights are not so clearly spelled out. Indeed, they may need to be negotiated from situation to situation. On the matter of how much alteration a participant or client may demand in a final report, serious differences of opinion between a client and an evaluator may occur. Madak (1994) writes:

The degree to which the evaluator is willing to make changes to the material, if it is objected to by a participant, is open to debate. That is, if the evaluator feels that she or he really heard or saw what was reported, then the text should not be altered just to please the informant. Altering accurate information might change the outcome of the study and thus lead the evaluator to inaccurate conclusions. (pp. 6–7)

My own practice as evaluator is to change or remove information that turns out to be patently wrong or, less clear-cut, that offends the participant’s sensibility. I find that sometimes the reader objects to some minuscule detail in a way that I feel is out of all proportion to its importance in the study (in much the same way that I, as client, objected to the inclusion of the adjective “dire”), and I quickly remove the offender. Other sources of concern can often be mitigated through changes in wording or the addition of a clarifying or qualifying phrase. In this instance I, as client, made my suggestions to Lynne but left the final decision whether to make the changes up to her. As we discussed these issues, our debate was sometimes forceful, but I feel that we both came to understand the other’s posi-

tion better, and that the report is stronger as a result. However, if I, the client, were a more forceful debater or Lynne a weaker evaluator, would this process have compromised the evaluation? There is no easy answer to this question. As House (1993) says, talking about evaluation of government programs:

Evaluators themselves are professionals and subject to the same pressures as other professionals working for the government, even though their peculiar task may be to evaluate other professionals. What should evaluators do under these circumstances? To what degree must they follow government policy and to what degree embrace the standards of their own profession? Ultimately, these are matters of professional ethics and social justice. (p. 55)

The Evaluator, the Client, and Utilization

Even though the original intention was for the evaluation to be formative in nature, in the end the report had a summative flavor. In retrospect, I believe that this was a design flaw. It would have been much more productive if Lynne had met with stakeholder groups to share her findings and at the same time work with them to see what could be improved. As it is, the substantial report will probably be seen as a summative product. This does not do justice, in my mind, to the early stage of implementation that the program has reached, because only one intake of the two-year post-academic stream has graduated, and only the pilot group of the four-year direct-entry stream has finished. Thus, frustrations of faculty and students are at a maximum and satisfactions at a minimum. Passing through such a slough of despond is probably inevitable when a program undertakes such a massive revision as this one did. Greenwood and Shanks (1996) report an uncannily similar pattern in their description of change in another teacher education program. However, there is no use crying over that particular issue, and the task at hand is to maximize the utilization of the report. That will be the client's job, as there is no expectation that the evaluator will take any further part in the proceedings.

Even if evaluators are to participate in the utilization of findings, they must appreciate that the fruits of their single-minded efforts are but one ingredient of the decision maker's soup. Rossi and Freeman (1989) try to temper the evaluator's missionary zeal:

First, evaluators must accept the fact that their efforts are but one input into the complex mosaic from which decisions and actions eventuate. Second, there are invariably strains that result from the conflicts in the interests of these stakeholders. (p. 424)

Stakeholder interests generate an emotional response to evaluation. Certainly, I had a strong emotional reaction to the report, although intellectually I am committed to using it in the best way I can. Scriven (1993) emphasizes the importance of understanding the depth of stakeholders' emotional response to evaluation, although he paints the issue with too broad a brush:

The roots of resistance to evaluation go very deep indeed, and it is wise not to underestimate their strength. For many people, to concede that their work needs evaluation is to concede that they lack competence. (p. 89)

This, I believe, was not my situation. Welch and Sternhagen's (1991) discussion of the emotional reaction to evaluation reports is more insightful:

Although anxiety is frequently suggested as a possible consequence of evaluation, the exact genesis of this type of anxiety is rarely considered ... Two of Erikson's five types of fears seem relevant to program evaluation. A program evaluation could result in either "intolerance of being manipulated" or "fear of losing autonomy." (pp. 123-124)

Of these two fears, fear of losing autonomy concerns me more. In the body of this paper, I reported how I worried that the report could cause the college to lose autonomy, or even cease to exist, a point made by House (1993). This fear remains. Another emotional response not mentioned by Welch and Sternhagen was the sense of being totally overwhelmed by the torrent of negativity. This response has abated for two main reasons. First, I have realized that the state of the new program and the morale of faculty are not exclusively my problems. In a collegial organization, no one person can fix problems. The situation is going to call for a concerted effort by all faculty members, and I am more than willing to play my part. Second, I have come to the conclusion that the situation is not as black as Lynne painted it. My own experiences as a faculty member and the

knowledge gleaned from my two case studies have bolstered me in this belief. As Thesis #94 (Cronbach and Associates, 1980) states:

Those who shape policy should reach decisions with their eyes open. It is the evaluator's task to illuminate the situation, not to dictate the decision. (p. 11)

But the evaluator can help the client toward action to improve the program. Securing the utilization of a report is a political business (Patton, 1986). Perhaps we would not always want to go as far as Popham (1995) advocates, but he does emphasize the need for persuasion:

Unlike strategies based on an advocacy of evaluation because "it is the right thing to do" or because "children will benefit," I contend that *we must start to sell educational evaluation because of its likelihood to directly and personally benefit the individuals who authorize it.* (p. 271; italics in original)

For Popham, this philosophy translates into action in this way:

We act when we see that we are *personally* apt to benefit. What I am recommending, therefore, is that educational evaluators identify the individuals who must authorize an educational evaluation, then persuade those authorizers that the conduct of the educational evaluation will, in a direct manner, preserve and enhance the manner in which the authorizer is regarded by others. (p. 271; italics in original)

It is sometimes hard for me to detect the personal benefit that will accrue as a result of this evaluation. But I don't think that I am being starry-eyed when I say that because much of my identity is tied up in being a teacher educator, helping to produce the best possible program for our teachers will indeed result in personal satisfaction.

REFERENCES

- Alkin, M.C., & Associates. (1985). *A guide for evaluation decision makers*. Beverly Hills: Sage.

- Braskamp, L.A., Brandenburg, D.C., & Ory, J.C. (1987). Lessons about clients' expectations. *New Directions for Program Evaluation*, 56, 63–74.
- Cronbach, L.J., & Associates. (1981). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.
- Greenwood, T.J., & Shanks, J. (1996). *Restructuring, learning and culture: Understanding the process of change*. Paper presented at American Educational Research Association National Conference, New York.
- Guba, E.G. (1975). Problems in utilizing the results of evaluation. *Journal of Research and Development in Education*, 8(3), 42–54.
- House, E.R. (1993). *Professional evaluation: Social impact and political consequences*. Newbury Park, CA: Sage.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage.
- Madak, P.R. (1994). Ethical considerations when using qualitative methods in evaluation. *Canadian Journal of Program Evaluation*, 9(2), 1–13.
- Meyers, W.R. (1981). *The evaluation enterprise*. San Francisco: Jossey-Bass.
- Patton, M.Q. (1986). *Utilization-focused evaluation*. (2nd ed.). Newbury Park, CA: Sage.
- Popham, W.J. (1995). An extinction-retardation strategy for educational evaluators. *Evaluation Practice*, 16(3), 267–273.
- Rossi, P.H., & Freeman, H.F. (1989). *Evaluation: A systematic approach* (4th ed.). Newbury Park, CA: Sage.
- Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (AERA Monograph on Curriculum Evaluation, No. 1) (pp. 39–83). Chicago: Rand McNally.
- Scriven, M. (1993). *Hard-won lessons in program evaluation* (New Directions for Program Evaluation, No. 58). San Francisco: Jossey-Bass.

- Watson, K. (1987). Contracting for program evaluation resources. *Canadian Journal of Program Evaluation*, 2(2), 81–84.
- Welch, W.W., & Sternhagen, F. (1991). Unintended effects of program evaluation. *Evaluation Practice*, 12(2), 121–129.

DEGREES OF STRANGENESS

Lynne E. Baillie
University of Saskatchewan

Abstract: Entering an alien environment and making observations from the perspective of the clear-eyed outsider is, perhaps, the most commonly understood role of the evaluator. Even when in a known setting, the evaluator is urged to “make the familiar strange.” However, there are occasions when that which was thought to be familiar, and that which was assumed would be strange, meet and merge in a manner beyond the control of the evaluator. The University of Saskatchewan College of Education, where the author carried out her graduate studies and taught as a sessional lecturer for several years, hired her to evaluate their recently implemented undergraduate program. In this role, she found herself employed by her doctoral advisor to interview and evaluate those who had been her professors and colleagues. This article reflects on that experience.

Résumé: Pénétrer dans un environnement inconnu et observer avec une perspective impartiale sont, peut-être, les rôles les plus compris d'un évaluateur. Même dans une situation connue, l'évaluateur doit la transformer et la rendre étrangère. Il existe, quand même, des occasions où ce qui nous est familier et ce que l'on assume inconnu se rencontrent et se mêlent d'une façon hors du contrôle de l'évaluateur. Le Collège d'Éducation de l'Université de Saskatchewan, où l'auteure a fait des études graduées et a aussi enseigné pour plusieurs années, l'a engagée pour évaluer le nouveau programme d'étude sous-graduée. Dans ce rôle, elle s'est retrouvée employée par son directeur de doctorat pour interviewer et évaluer ceux qui furent ses professeurs et collègues. Cette présentation est sa réflexion sur cette expérience d'un retour “à la maison” pour évaluer une institution en voie de transition.

In September 1995 I was invited by the Assistant Dean of Undergraduate Programs to submit a proposal for an evaluation of the

undergraduate program in the College of Education at the University of Saskatchewan.

Initially, I was reluctant to carry out this evaluation. From my years in the college — as a graduate student, a sessional lecturer, and a college supervisor — I knew the people who would become my participants. What I was *not* familiar with was this entity called the “new program.” I had been involved in some aspects of the piloting of one core course, but the program as a whole had been implemented in the three years following my departure from the college. I had kept in touch with friends who were involved with the implementation of the program during this time, and so I knew of some of the delivery complications. From my position on the periphery, it did not sound as if it had been an easy birth. The thought of conducting an evaluation of this new program led me to fear both what I might find and what I might *not* find.

I also had other concerns: I was unsure whether I could successfully negotiate the political minefields of the college departments and administrative structures and emerge with a credible evaluation; I felt that this evaluation was long overdue — that a more formative and integral design involving ongoing dialogue with program participants would have been more appropriate; I wondered if I could be fair and unbiased knowing the possible consequences of reporting negatively on the program in the present climate of funding cut-backs and falling student enrollments. In short, I questioned whether I could — or wanted to — undertake such a task, and do it passionately and honestly.

I arrived at the decision to submit my proposal to the Undergraduate Program Committee (UPC) for many reasons. From my previously mentioned vantage points of graduate student, lecturer, and supervisor, I had recognized in some of those I worked with, and learned from, a professionalism, a commitment, and an indefatigable drive to do the best job they could as they worked to transform students into teachers. I wanted to understand what it was about the new program that had made these individuals begin to question both what they were doing and why they were doing it. I wanted to know how an institution with a long-established and primarily technical-behaviorist approach to teacher preparation was rebuilding itself to comply with the transformative-constructivist vision contained in the new program. I also had great confidence in the “client” with whom I would be working: Alan had been my doctoral advisor, and we had since worked together on several other projects

and evaluations. From our previous joint research projects, I knew Alan to be a well-informed and scrupulously honest research partner. I could look forward to the usually animated and always demanding and stimulating discussions that had become part of our collaborations.

Taken together, these reasons outweighed my doubts, my proposal was accepted, and I duly began my research into the new undergraduate program in November 1995.

DANCING IN THE MINEFIELDS: OPINIONS, "TRUTHS," AND PURPOSE

The first indication I had that this evaluation was to be more than a measure of program strengths and weaknesses was the immediate reaction from those delivering the program. In my proposal, I had stated that I would restrict my interviewing to selected and representative stakeholders — program designers, program administrators, faculty, students, and cooperating teachers — via a total of 18 individual interviews and 3 group discussions. It is indicative of the general anxiety and concern felt over the new program that once people became aware of my appointment, *they* contacted *me* — wanting to talk, share, inform. In almost every case, their need to be heard was unmistakable. It seemed to me that many were interpreting the evaluation in terms of the program process itself, and not as a discrete occurrence. For some of those who, for whatever reason, had felt left out of program design and delivery decisions, inclusion in the evaluation took on a deeper significance than just a review of the program would warrant. Failure to include these individuals, I believed, not only could damage the the credibility of the final evaluation report in their eyes, but also could further increase the already considerable distance they felt from the program. With these possibilities in mind, I decided to extend my interviewing to include as many of those involved in the new program as possible. This brought the total to 59 individual interviews and 7 student group discussions.

The second indication that this evaluation could contain other meanings and purposes came via some of the information presented to me during interviewing, and the manner in which that information was delivered. This was also my first insight into the shape-shifting that was to become a feature of the evaluation, as roles and relationships lost their definition and form. Listening to evaluation par-

ticipants, I knew that I was being told what they wanted me to hear. I don't think that this is at all unusual in data collection. However, what complicated the issue was that I was unsure of the ways in which the *function* and the *perception* of role were played out in our interactions. Role affects perspectives, for as much as who we are, what we feel we should project influences how we categorize others and their actions. If I was, indeed, being manipulated on occasion, then in which role were participants doing the manipulating, and how was I being perceived in the process? How did the people I was interviewing interpret the relationship between us? Was I graduate student and sessional lecturer still? Did some participants retain a degree of professorial authority over my actions? Rather than manipulation, was this a misplaced, albeit well-intentioned, guiding toward what they thought I should be doing? Although I believe the majority of my interactions with faculty to have been entirely open conversations about working within the new program, I occasionally found myself having to decide whether I was hearing direct and frank discussion because I had absolutely no place in the hierarchy of the college or, more disturbingly, if I was being fed some juicy morsel because I was perceived as being a naive and conveniently direct channel to UPC, and was not being seriously considered in the role of evaluator.

These indications came early in the process of data collections and made me very cautious about what I was hearing and what *my* role was in this evaluation. As the evaluation progressed and the negative content of participant responses accumulated, my caution increased. The product of the evaluation process is information, and, as we all know, information readily equates with power. Careful consideration of whose information should be included, and for what purpose, guides every evaluation. However, in this instance I found that I was faced with a confusion of expectations, roles, and purposes and was overwhelmed by diverse "truths," opinions, and purposes. I became almost obsessed that, unwittingly, I would reify political stances at the expense of understanding; that I would fall into the trap of attempting to find answers without first framing the right questions; that I would allow one truth, opinion, or purpose to become the controlling factor in the evaluation.

THE VIEW FROM HERE

Alan and I did not escape from this dilemma. We too became part of that multiplicity of perspective and role as we stated our beliefs and

defended them. Some examples:

- A semantic debate over a single adjective, ostensibly involving client/evaluator confirmation, seemed to me to be about much more. I felt that it was really about how prepared we were to deliver what had become a negative report. It was about how much we trusted each other to do what was the right thing under those circumstances, and indeed, whether we had a common understanding of what the “right thing” meant. The adjective was removed.
- We discussed, on more than one occasion, whether or not to include quotes from a particular participant. The view from Alan’s side of the desk was that this individual was mistaken — that despite being given the correct information time and again, this person still clung to an idiosyncratic and faulty interpretation. The perspective from my side of the desk was that this person, wrong or not, was functioning within the new program — and the college — according to that interpretation. This, I thought, must surely affect how the program is being viewed and implemented in that particular department. The quotes remained.

Here again we see the influence of roles. In the first case, I used the contentious adjective because, as a graduate, I was devastated at what I had found happening in my “home” institution. As an administrator, Alan saw the word as he would have to explain it to the media. I was seeing a quote as information-to-be-included, with the immediacy of the evaluator. Alan was seeing its inclusion from the more pragmatic stance of the person who would have to make decisions and then implement them.

Sessions like these invariably led to discussions of the limits to the roles of evaluator and client. Although I do not believe we ever arrived at any fixed definition of those roles, the discussions did raise questions concerning our current situation. We both came to this evaluation with the memories of work we had collaborated on in the past. When it became apparent that the current experience was to provide more differences than similarities, the “old” system partially broke down. Until we established new and mutual reassurances we were, for an interlude, unsure of just how to approach each other. Undergoing this change was, I think, an uncomfortable time for both of us.

Change and the Power of Remembering

We may have fairly good control over what we think, but we are much less masters of how we think. We may see the world clearly through the lenses of our eyeglasses; but if we have worn them for a long time, we often forget we have them on. A refracted view is how we come to think things really look. (Kantrow, 1987, p. 25)

My involvement in this evaluation served to remind me that *real* change is uncomfortable and hard. To change involves stepping away from the solid and familiar onto shadowed and uncertain ground. Real change requires a completely different way of looking at the world, a shift of vantage point that affords a view that may not be entirely new but has, up until now, been hidden. If it has also gone unsuspected, then the new perspective will cause some considerable discomfort, as the old maps that accompanied the familiar view will require revision. Consequently, too much change can be overwhelming and is actively resisted by organisms, such as ourselves, that inherently seek security and stasis. The feeling of strangeness, the initial lack of fit, that comes with newly defined roles and external expectations is not a welcome sensation within an established institution, and it is all too easy to take refuge in self-protective behaviors while clinging to the familiar.

Change and the Institution

Kantrow (1987) writes that institutions also remember and seek stasis:

Like it or not, the past infects the world we live in, the decisions we make, the choices we see before us. If we ignore its influence, we do not escape its power. All we do is remain to some extent its prisoner without ever really knowing that that is what we are. (p. xiii)

As my analysis of data proceeded, I turned my attention to the wider institutional context of the new program, not just the program itself. I believed that by doing so I could expose some of the constraints that were clearly working against program success. The most influential of these was the institutional barrier of tradition.

The College of Education carries the legacy of its positivist history. The program that preceded the new one was predominantly techni-

cal and prescriptive in its outlook. Dissatisfaction with this approach motivated the revisions that led to the implementation of the new program. The philosophy informing that program is one that reconceptualizes teaching as informed problem solving and that includes consideration of the moral aspects of teaching. The central tenets of the new program are reflective practice and critical decision making. In attempting to accommodate the demands of the new program while remaining influenced by a powerful institutional memory, the College of Education finds itself caught between two quite distinct paradigmatic discourses of teacher education — the normative and the dialogical (Britzman, 1991).

The college has delineated the categories and criteria for the work of teacher preparation. Over time, these have come to be objectified and are seen by many as being natural and unchangeable. But Beauchamp and Parsons (1989) remind us that all programs and curricula

represent ideological and cultural biases which come from somewhere. Conceptions of student teacher competence, good performance, and proper behaviour are not free-floating ideas; each is a construct laden with values. (p. 125)

The fact that these categories are creations is forgotten or overlooked; they are rarely if ever questioned. And because the technological metaphor is embedded in the college structure, there remains an unchallenged legacy that is at odds with the vision surrounding the new program. For example, a central intent of the new program is to create an environment where students can critically reflect upon their personal practice and upon curricular decisions. Within the unexamined context of the institution, this environment failed to materialize, not because people do not want it to happen — although that can also be understood as a legacy from the past — but because the attitudes, traditions, and expectations that existed long before the new program came about have created a situation where there is simply no time or place for it to happen.

It appeared to me that the people I interviewed were reacting; although they had been invited and encouraged to contribute to the design of the new program, it seemed that many of them knew what they wanted only after they saw what they got. How can these institutional barriers be overcome so that people can communicate ideas

and visions beyond the collective political gamesmanship and the shaping influences of history and tradition? How many “degrees of strangeness” can be tolerated before resistance occurs?

Discussing the Undiscussable

Argyris (1990) defines the “undiscussable” as group actions or policies that prevent individuals or groups from identifying and resolving the real problems besetting the organization or institution. These organizational defensive routines are protected and reinforced by everyone within the organization who denies that they too have a role in such routines. Individuals do not like to take responsibility for creating or maintaining defensive routines. They can recognize the power of those routines, and in many cases are willing to say that they are personally influenced by them. However, they are unwilling or unable to admit their own contribution to the creation or maintenance of such routines.

The most common reaction noted by Argyris, and one that recurred in participants’ responses in this evaluation, is that of cynicism: a perspective of “nothing will ever change around here, and there’s no point in trying” or “no-one ever really listens, so why speak out.” If other people are perceived as being the cause of the difficulties experienced, it becomes easy to assume an attitude of helplessness and defeat. In this way, individuals learn to fault the organization while not feeling personally responsible for correcting the faults. Responsibility becomes a problem for administrators; faults in the system, by the same logic, are then seen as attributable to poor administrative practices. Individual members of the institution’s dissociation from responsibility can therefore encourage mediocrity of performance and lack of commitment to engage in any process of change. There is consequently an inevitability, a kind of self-perpetuating logic, to defensive routines.

Individuals feel helpless about changing organizational defensive routines. They feel the change is hopeless, because the cure appears to be one that will make the illness worse. They do not wish to be seen as deliberately making the situation worse by opening up a can of worms. The result is something equivalent to an underground economy — a grey organization that is alive and flourishing, yet officially dead or non-existent. This, of course, makes it possible for the grey organization to

remain alive and to flourish. We now have the underground managing the above ground. (Argyris, 1990, p. 43)

The central challenge facing the new program at the College of Education is the context in which it has been placed. It has already been drastically altered and its possibilities diluted by existing forces. These forces are predominantly institutional in nature and should not be mislabelled, and consequently treated, as fiscal or demographic.

Organizations invariably respond to factors other than the attainment of their formal goals — the salience of the drive for organizational perpetuation, personnel's need for status and esteem and their attachment to the practice skills in which they have invested a professional lifetime, conservation and inertia and fear of the unknown consequences of change, sensitivity to the reactions of various publics, costs, prevailing ideological doctrines, political feasibility, etc. (Weiss, 1972, p. 319)

This suggests that it is not the new program that will determine the success or failure of the college, but the role and function that the college defines for itself that will decide the success or failure of the program. Without accurate identification of the role and function of the institution, any modifications made to the new program will be exposed to the same institutional forces that restructured the original vision. This cycle is capable of endless repetition (Aoki, 1991).

It seems that change in an existing institution cannot be brought about by a single stroke. Decapitation may certainly cure the patient of migraine, but the general prognosis is less than desirable. Institutions comprise a complicated and organic weaving of individuals, groups, roles, and functions, each with its own threshold of tolerance for change, growth, and difference. Moving away from the old toward the new should, therefore, be a slow and systemic process, with attention being paid to the organizational maintenance imperatives that influence and restrict attempts at change, and the powerful seduction of familiarity.

The "new" program had been running for three years before this evaluation was initiated, and although one of the more obvious restrictions placed upon this evaluation was its summative nature, I remain uncertain as to whether a more formative approach would have produced different results. In fact, the active and continual

involvement of the major players in the new program may have confined the evaluation to the new program itself, and consequently obscured the powerful and distorting effects of wider institutional forces.

The findings of this evaluation lead me to believe that the real potential of the new program at the College of Education lies not in terms of any inherent innovation, but in the fact that its implementation throws institutional difficulties, accrued over some 20 years, into undeniable and insistent relief. Within that understanding, the evaluation becomes not only a collection of findings and recommendations for change, but part of a process that Mayne (1992) refers to as “organizational learning,” providing a platform where accepted institutional wisdom can be questioned and challenged.

When considering change within established and traditional institutions, I see identifying and communicating old constraints as more critical than — and as preceding — the development and implementation of new initiatives (Conrad, 1990). Unless the power of these established and taken-for-granted ways is recognized, the new program — *any* new program — will in time become a mere approximation of everything that has gone before.

CONCLUSION

We all hold precious beliefs that we nurture and guard closely, sometimes despite all evidence to the contrary. Even beyond political maneuvering for personal or group gain, individuals will naturally protect the opinions they value and the territory where they feel safe. I found that to include this reasoning in an evaluation is not easy. Attempting to achieve a fair representation that includes multiple perspectives is like trying to create an accurate image while walking down a corridor of carnival mirrors. But it must be attempted. The information presented to the evaluator has to be understood as being relative to the perceiving participant first, and then in terms of the evaluation purpose. Even if the intent of the information is political, it is not for me, as the researcher, to decide that that participant should be silenced, for then I would have abandoned the evaluator’s role in favor of that of judge. I can provide the contexts for the words of my participants, listen for the repetition of particular stories, and try to present my findings in a way that invites the critical response of the reader (Gitlin, 1990). I cannot editorialize. Similarly, I cannot categorically state to readers that

this, what I report, is reality, for whose reality would I be reporting? I can only say that *this* is what I found; *this* is what the people who work here say and do; *these* are their perceptions and this is what appears to be happening.

The evaluation is now completed and submitted. Questions of interpretation are for the readers of the report to answer. I too have questions. The processes of this evaluation have left me with the seeds of new thinking; I found probing the grey areas that exist between program aspirations and institutional realities to be troubling and, at times, quite painful, leaving me with changing views of both the function of evaluation and my role as an evaluator. As yet, many of my questions remain unanswered, but perhaps, like the transitional institution, I need to move slowly and with caution into this reconsidered landscape.

REFERENCES

- Aoki, E. (1991). *Inspiriting curriculum and pedagogy: Talks to teachers* (Occasional Paper on Curriculum Praxis). Edmonton: Department of Secondary Education, University of Alberta.
- Argyris, A. (1990). *Overcoming organizational defenses*. Englewood Cliffs, NJ: Prentice-Hall.
- Beauchamp, L., & Parsons, L. (1989). The curriculum and student teacher evaluation. *An Interdisciplinary Journal of Curriculum Studies*, 9(1), 125–171.
- Britzman, D. (1991). *Practice makes practice*. New York: State University of New York Press.
- Conrad, C. (1990). *Strategic organizational communication*. Fort Worth, TX: Holt, Reinhart & Winston.
- Gitlin, J. (1990). Understanding teachers dialogically. *Teachers' College Record*, 91(4), 537–562.
- Hudson, J., Mayne, J., & Thomlinson, R. (Eds.). (1992). *Action-oriented evaluation in organizations: Canadian practices*. Toronto: Wall & Emerson.
- Kantrow, A. (1987). *The constraints of corporate tradition: Doing the correct thing, not just what the past dictates*. New York: Harper & Row.

- Mayne, J. (1992). Enhancing evaluation use. In J. Hudson, J. Mayne, & R. Thomlison (Eds.), *Action-oriented evaluation in organizations: Canadian practices* (pp. 285–287). Toronto: Wall & Emerson.
- Weiss, C. (1972). *Evaluating action programs: Readings in social action and education*. Boston: Allyn & Bacon.

EPILOGUE TO:

THROUGH THE LOOKING GLASS: WHAT HAPPENS WHEN AN EVALUATOR'S PROGRAM IS EVALUATED *and* DEGREES OF STRANGENESS

Alan G. Ryan and Lynne E. Baillie

Evaluation is not a bloodless undertaking. It involves the interaction of real people who see the world from different mountain peaks and who have interests they need to protect. The experience described in the preceding two articles has allowed us both the rare opportunity to step out of our usual roles and to experience the roles of others. Dealing with the consequences of conducting the evaluation required a difficult and painful re-examination of personal functions and reflection on concomitant attitudes, approaches, and interpretations. Neither of us was entirely comfortable with the findings of the evaluation. The steps that were taken — and continue to be taken — in order to accommodate the outcomes of the evaluation are, at an individual level, a reflection of the greater adaptation to change that faces the collective institution.

For those who like to know the end of the story, here is a brief account of how the report was used. Members of the Undergraduate Program Committee read and considered it carefully. They made a crucial distinction between *program* issues and *personnel* ones and decided to focus on the program concerns, as that is the committee's mandate. The personnel issues were left to the incoming dean to handle. Although the report was never released in full, detailed extracts from it were used in a subsequent committee report that outlined its recommendations for changes to the program.

The experience has changed and broadened both of us. The “evaluator” has explored the uncharted terrain lying between program aspirations and institutional realities: What is it reasonable to expect of real people in real situations? How does the evaluator reconcile dreams with constraints? The “client” has gained a deeper appreciation of the complexity of that role and a humbling understanding that an evaluation’s findings are but one dimension of the client’s reality.