

The ‘presence’ of evaluation theory and practice in educational and social development: toward an inclusive approach

Murray Saunders*
Lancaster University, UK

This paper outlines a vision of evaluation and its place in social and educational policy and practice. It focuses on the ‘presence’ of evaluation in theory, organizational learning and internationalization and the ‘voice’ of participants in the evaluation process drawing on a range of examples of evaluation practice. It argues for an ‘inclusive’ evaluation stance from a moral/political standpoint and from the standpoint of sound evaluation design. It offers evaluation as a way of promoting and depicting the effects of social policy on its recipients and concludes by suggesting the way evaluations can promote ‘provisional stabilities’ for those experiencing rapid and complex change.

Introduction

Evaluation can be a frustrating business; one major source of frustration is its potential ineffectuality. That is to say it often fails to enter the decision making process at the right time, saying useful things to the right people such that positive contributions to development can be made that do justice to the complex and contradictory experience change can involve.

This paper depicts the ‘presence’ of evaluation in social and educational policy and practice and explores how it is made apparent, or, in many cases, has stubborn invisibility. The word ‘presence’ is used as a way of capturing evaluative practice as we move into the first decades of the twenty-first century. The idea of presence is not neutral for me. Evaluation is a means of giving voice to the disadvantaged, dispossessed or disenfranchised and their actual and potential relationship with policy implementation, but also concerns the unacknowledged, the unanticipated and the unintended dimensions of evaluation practice. As well as ‘presence’ I would

* Centre for the Study of Education and Training, Department of Educational Research, Lancaster University, Lancaster, LA1 4YD. Email: m.saunders@lancaster.ac.uk

also like to use the metaphor of ‘voice’ when referring to people or groups with a latent or explicit interest in the evaluation process. Voice has a declamatory meaning, but more prosaically, it can simply refer to the interests stakeholders have or potentially have in the evaluative process.

In this way, I would like to focus on five dimensions of the presence of evaluation. These dimensions are personal in that others might have a different list but they are not arbitrary. Like many evaluators, I fell into evaluation some 25 years ago and began to do it before having real awareness of a body of knowledge associated with evaluation, let alone a ‘community of practice’ or more strictly speaking ‘communities’ of practice that had a professional self-consciousness. It is on this experience that I draw in this paper, making it a kind of manifesto or at least a declaration of what seems important to me in professional evaluations.

The dimensions are:

- The voices of the ‘recipients’ of evaluation programmes and projects.
- The voices of the potential users of evaluation.
- The presence of analogous processes in institutional and social learning.
- The presence of theory in evaluation.
- The international presence of evaluation.

The paper suggests that the potential of an evaluative presence in all these situations, from participants, users, theorizers and politicians’ points of view, often struggles to be imagined, let alone manifest. This paper will draw on experiences of ‘doing’ evaluation in diverse cultural situations but mainly in the worlds of education or work. Evaluation is taken to mean the purposeful gathering, analysis and discussion of evidence from relevant sources about the quality, worth and impact of provision, development or policy.

In social policy areas like education, social services, criminal justice and health it is possible to identify four clusters of evaluation practice. These are identified in Figure 1.

- **Systemic** (embedded processes of inspection and regulation often in the form of quality assurance processes, ranking and auditing within a social policy sector)
 - **Programme** (specific evaluations, usually external and for which individuals or groups have competitively tendered and which are associated with a particular intervention, policy or programme)
 - **Internal** (undertaken within an organisation, usually top down to make judgements about the performance or quality of specific parts of the organisation, often conflated with self evaluation)
 - **Self** (initiated and undertaken by a group of practitioners for the purpose of problem solving, development or improvement of their own practice)

Figure 1. Clusters of evaluative practice

The paper discusses mainly the presence of evaluation in the evaluation of programmes, policies or social interventions illustrated in particular by reference to evaluations undertaken in CSET (Centre for the Study of Education and Training,) in the Department of Educational Research, Lancaster University, UK.

The programme recipients

Programme recipients are sometimes confused with the 'evaluated'. We can identify the 'evaluated', as the initiators of policy or programmes (it is their intentions after all that we might be evaluating). However, often, in the minds of those involved in evaluation as commissioners (usually funders of programmes), the 'evaluated' are the recipients of policy, i.e., those that provide the main source of evidence of its efficacy or, put another way, those who will experience the programme's effects, good or bad. This is of course misleading. Sometimes, this latter group is called, rather clumsily in my view, 'evaluands'. In an important sense we are in a situation in which the programme's 'target' or 'recipient' group constitute the main source of evidence for evaluative judgements, they should not be confused with 'the evaluated'. For the purposes of this paper, I will call this 'target group', 'programme recipients'.

I would argue that the voice of programme recipients should be fore-grounded in evaluation design on the basis that their voice will authenticate and validate the provenance of the evaluation and improve and strengthen its design. Their experience should be articulated faithfully by the evaluation and it is on this basis that the evaluated, that is the policy-makers and programme designers who promulgate policy, will have the best resources on which to make judgements about their policies or programmes.

The rest of this section of this paper considers the recipients of programmes or the group at whom a programme might be targeted or more optimistically, with whom a programme is being developed. To that extent it constitutes an inclusive evaluation declaration.

Their voice can be heard in four ways:

1. By involving them in identifying and using key questions, indicators or issues (concern with participatory approaches); outlined graphically in empowerment evaluation) at a 'strong' end of the participatory evaluation continuum (see Fetterman *et al.*, 1996, and its critique by Patton, 1997).
2. Being part of an ethically justifiable process (a concern with evaluation ethics).
3. Making sure their experience is faithfully reported even under political pressure (a concern with declamatory platforms).
4. Evaluation products entry into a public debate (a concern with evaluation as part of a democratic process and as a way of promoting democratic participation).

The first characteristic of this inclusive approach is authentication by simply asking the programme recipients to identify what the key questions might be that cut to the essence of a programme's effects on them. This differs from Patton's

(1996) conception in that he emphasizes the *commissioner's* role in design. In my schema this is more likely to be 'the evaluated' or promulgators of a programme than the 'recipients'. The process of involving recipients can be achieved through workshop designs in which representatives of the target group are provided with an opportunity to present the questions or indicators that will, in their experience, yield a good depiction of what may happen to them or will potentially happen. Undertaking this process at the outset of an evaluation also has a declamatory dimension. There is potential that this group's interests in the programme are embedded into the evaluation design.

This technique can be used successfully with very different target groups in our experience, with children aged 12-years-old in Mexico City (see Saunders, 2001a) and experienced academics in higher education in Scotland (see Saunders *et al.*, 2004). The central issue here is to ask recipient's to 'rehearse' the practical realities of the policy or programme and ask deceptively simple questions (which often yield complex answers): if this programme is to be a 'good' thing for you, what kind of experience will you have? Of course, the answers to such questions may or may not coincide with the programme designer's or evaluation commissioner's view of an indicator of success, however, once this process is undertaken, this authentic voice can be included as a design principle. In most cases, recipients are the programme's harshest judges so this is no refuge for the soft or banal indicator. The challenge here is the level of programme knowledge a recipient might have.

The young people in Mexico City (see above), and their experience of a Children's Rights Citizenship Project is a case in point. Part of the project had involved a virtual partnership arrangement with UK schools via email. In many instances, the contact with the UK schools had been sporadic or in some cases non-existent. While there were good reasons for this and the overall experience of the intervention from the recipient's perspective had been a positive one, the quality of the contact with the UK schools was identified by them as a very important indicator for the project.

The second characteristic is through agreements on ethical procedures associated with evaluation process. Many evaluation societies have produced guidelines on ethics that emphasize the interactions between stakeholders in an evaluation that express fairness, appropriateness, rights and obligations (see for example those produced by the UK Evaluation Society). I have observed, however, that the less power a 'recipient group' might have, the less access they have to the power of veto, access to evidence gained from them to check accuracy, capacity to exercise worries over anonymity and confidentiality. This puts a weighty obligation on evaluators to attend to their interests and to listen to their voice. It is routine for example for 'key informants' in an evaluation, usually people in powerful or influential positions, to have a lot of control over what goes into the public domain and access to reports and evaluation outputs. The same cannot be said of most participants in the evaluation process. More involvement can present logistical problems for an evaluation but involves recipients having sight of the data they have provided to check on accuracy and to involve recipients in building interpretations and theories about the data as

the evaluation progresses (for examples of this process see Saunders *et al.*, 2005). Dissemination of drafts as well as finished texts to recipients through active workshops involve recipients more realistically.

The third characteristic is embedded in the way in which recipients' experiences are accurately represented or depicted through evaluation reports and feedback. This is the most declamatory of an inclusive evaluation's aims. Unfortunately, however, depictions have not always had a noble history and data freely given can return to wreak havoc. In extreme cases, evaluations have resulted in terrible abuse. The socio economic evaluations of rural Vietnam which preceded the Vietnam War, for example, led to at best wide spread disruption and dislocation of rural life and in some cases and at worse to the annihilation of whole communities (for this culture clash see Fitzgerald, 1972). However, later analysts pointed to widespread ignorance of aspects of Vietnamese life that led to an underestimation of the Vietnamese will (see Spector, 1985). More commonplace is for evaluations to be used for rationalization, for forearming planners and policy-makers to potential areas of resistance which in turn provides opportunities to undermine or circumvent legitimate worries surfaced during an evaluation. The interesting comparisons made by the National Literacy Trust (2005) of Ofsted evaluations of the literacy strategy in primary schools and that of the external evaluations from Michael Fullan's team which point to the way Government tended to select out teachers' worries about over-ambitious targets and initiative fatigue, are cases in point.

More positively, however, unintended effects, legitimate concerns and opportunities to access power can be identified through an evaluation. Powerful accounts of ingenuity and creativity can be made available which can, on occasions, inspire others. An example of this is a women leaders project evaluation in Chile (see Saunders, 2001b) in which an intervention did provide a development opportunity which was captured and disseminated through an evaluation reporting process.

This evaluation was of the effects of several interventions (funded by the British Council) but specifically of a five day seminar on Women Leaders in NGOs (non-Governmental organizations) informed by the following situational analysis offered by a commissioner of the intervention:

Women have historically played and are playing a major role in NGO work. The objectives of the seminar were to strengthen the management and leadership capacities of women working in NGOs and to underline the gender issues involved in the way they work, in order to help them and their organization respond to new circumstances. Since the return of democracy, the situation of NGOs in Chile has changed drastically. On the one hand, there is no longer the common aim of fighting for a democratic Government, and, on the other, international agencies are not providing the kind of support they used to get during dictatorship. In this context, many NGOs are questioning their objectives and even their existence.

The seminar had a high profile, with the Minister for Women providing an opening address. Forty women from NGOs throughout the country participated, all had management responsibilities for their organizations at different levels. The content of the seminar encompassed two distinct strands:

- Leadership and management development.
- Gender related issues.

Participants were invited to a follow-up meeting with the aim of assessing the longer term impact of the event. The most significant impact of the seminar would be knock on or multiplier effects or activities prompted by the original experience. According to the central commissioning team, some women had led or were leading changes in the organizational culture of their NGO and many had organized training workshops for their female colleagues or customers based on the original seminar.

In order to triangulate these observations and to depict the positive effects of the original experience, a meeting was arranged with some key informants who were original participants in the workshop to discuss subsequent use of the workshop experience. Informants from an NGO working in a small town on the outskirts of Santiago were interviewed. The NGO had a wide-ranging brief to provide support and development to excluded groups (local women, disaffected youth, indigenous people) and specialized in vocational training, appropriate technology solutions and environmental awareness-raising.

The general experience of the event is captured in the following quote from one of the informants:

The level of participation was really good, there were lots of different people but they were either from NGOs or women who work for others. The seminar had a speaker for each topic and we formed groups to work on the topic, basically strategic planning from a women's point of view, to elaborate the strategic planning ideas we tried to apply them in small groups. The problem was that there was too little time although the speakers were very strict. There was too little time to discuss things in the groups, one more day would have made a big difference, also it is a shame that we could not all stay in the same hotel but we thought the seminar was wonderful, all the topics were useful for us. I think it was a great idea, so I am applying the seminar ideas with poor women in this area.

The original participants pulled together women from a variety of groups in the local area, each of which had rather different goals. They used the seminar ideas to mobilize the women to join an overarching group in order to improve their bargaining power to procure resources.

We did the same strategic planning within five months. At the end of this period we did the same kind of workshop in a district called Buin. All kinds of women came together to do some useful planning. The one thing they had in common was that they couldn't organize themselves, some couldn't write. But, there were some very practical outcomes with this group of women who came from different sectors. We basically presented the same programme. We asked them to check-why am I doing this? What I am doing now? What are my goals?

Some of the problems with this 'cascading' idea were in the area of 'transfer'.

There were many ideas which we could not achieve in practice, how to manage money for example. The main problem is that we work with rural women who are the poorest, may not know how to read, while the original women were very well educated. The main thing we wanted to get over was the idea of distinguishing between goals and the

means of achieving the goals. A big problem was to get real consensus between the women, we had to get consensus on the goals, it was very difficult because they didn't really have an analytical approach.

Perhaps the most important impact of the original intervention by the BC was in providing the initial impetus for this group of women to enter 'civic' society effectively and begin to organize politically and socially.

The greatest thing was helping them realize what they really wanted, to be organized themselves. There were 12 separate groups and they ended up as one unified group, they could now belong to the 'civic town'. Now they had an organized structure to obtain funds.

The informants were also able to identify a personal more individualized impact of this process in addition to the organizational as they state in their own words. This is a powerful 'recipient' endorsement of the relevance and potency of this activity choice by the project funders.

Another important thing is to try and change them into a political person from a person who did nothing of this kind-to fight for what they need. I suppose they were empowered. In two weeks they were presenting proposals. By the way, the unified group is called BUCAM (the communal union of women of Buin). This is very important politically, because they can go to the local community now they have a voice and we can reproduce the process in other areas in the suburbs

The socially positive potential afforded by evaluations does not always emerge. Evaluators as a group have uncertain power over evaluation use (I will explore this point in more detail below). In order to improve these possibilities, discussions over evaluation use that are built into designs at an early stage, including ways of representing unpalatable or inconvenient messages, address this issue. The presence of those who are the intended 'recipients' for social intervention programmes, often the only source of evaluation evidence, should be consciously built into evaluation designs. In doing so, not only will evaluations be better, more accurate and authentic accounts of experience, but the legitimate voice of this group of stakeholders can be depicted.

The fourth characteristic concerns the contribution to democratic impulses evaluation promises. There has been a long tradition in evaluation circles to situate the evaluative impulse amongst the 'good guys'. To be more explicit, to see evaluation as something that should be done democratically (see comments above on 'inclusive evaluation') and as something that contributes to democracy. I align myself with both these aspirations. From the UK evaluators like Barry MacDonald and Saville Kushner (2000) and from the US, strongly associated with the work of Ernie House (1998), we have expressions of the way evaluations have the potential to contribute to democracy through the provision of resources for public debate on policies and programmes. To some extent the first of these aspirations is addressed in the preceding paragraphs and refers to the ethical procedures I have outlined and the involvement of 'recipients' in design. A further democratic consideration concerns the public nature of evaluation outputs and the extent to

which the evaluations of publicly funded programmes should be in the public domain (see below). Generally speaking the writers cited above argue for a position of 'openness' and freedom of access to afford the public the same knowledge based privileges as commissioners and programme designers. Difficulties can arise where programmes are young or undeveloped and early exposure to negative or critical evaluations can be unfairly damaging. Overall though, the democratic gains to be had in opening up access to information on the way in which public money is spent should be weighed against the tendency to control discussion on policy implementation and limit access to evaluative evidence to the political elite. The idea of evaluation acting as a counter to centralism and control in governance is discussed in the last section of this paper.

Despite this health warning, public policy can be informed by evaluations. Recent developments in the area of higher education policy on teaching and learning (the Learning Teaching Support Network and the Higher Education Academy) is a case in which evaluations of the various strategies for support and improvement were externally evaluated and the evaluation reports were a central part of the policy development process (for an example see Saunders *et al.*, 2002). Some of this inclusion in policy-making is a matter of timing, in some cases a matter of friendly 'gate-keepers' and in others, simply a matter of confirming or cohering with existing policy inclinations. This said, the development of an 'evaluative culture' in which an informed and participative approach to depictions of policy experience is a worthy aspiration.

Uses and users

While commissioners and users of an evaluation are not synonymous, they can be collapsed for the purposes of this paper. It is self evident that once an evaluation enters the public domain, if it does, then anybody can be a potential user of the evaluation if they have access to it. This is one of evaluation's great potentialities and suggests to me the urgency of establishing the levels of public access to evaluations very early. Also the thrust of earlier points on inclusive approaches implies both process use (see Patton, 1998) and access to the products of an evaluation by a wide range of stakeholders is an aspiration. A general rule should be that if the evaluation is in the realm of public policy, promulgated by public policy-makers, then the evaluation findings should find their way into the public domain.

There has long been a debate within evaluation circles about the relationship between an evaluation product (report, analysis, synthesis, etc) and the way an evaluation might be used. Should an evaluation contain recommendations for example? The position that evaluations should not make recommendations in the strictest sense of the word is persuasive. Ernie House (1998) argues that recommendations spill out of the legitimate purview of the evaluator into the realm of political decision making. He even goes so far as to say that it is an abuse of the evaluators' position to provide recommendations. This does not mean that evaluators duck the responsibilities I have been at pains to outline above. In order to come off the fence for an

inclusive stance, it is crucial to engage with decision makers. A recommendation is precisely that, i.e., a suggestion that the decision maker should do something specific on the basis of the evaluators' analysis. How can evaluators be in a position to do this? What they can do is engage in a 'conversation' that might outline options according to different stakeholders' experiences, or against stated objectives where we might be able to infer likely gaps, etc. We can also engage with decision makers as participants in the decision making process as evidence is discussed and implications for policy-making are reviewed. None of these things are the same as offering recommendations. It is also possible to check on the way evaluation evidence is used and presented, to offer a critique when it is distorted or used selectively, to point out the effects certain policy directions might have on what the evaluator knows about a stake holding group.

Recommendations have embedded within them a notional decision about what to do, otherwise they would not be a recommendation. This power legitimately belongs to policy-makers or commissioners. The role of the evaluation is to provide resources on which these decisions might be made but not to make the decisions themselves or to imply a 'latent' decision. This is not just a hair splitting exercise. Evaluations are constantly put under pressure by commissioners and designers of programmes to make recommendations. Sometimes this pressure is designed as a way to avoid any negative political fall out arising from evaluation implications and displace responsibility to the evaluator. More often than not it is simply a matter of not making clear at the outset, what the limits and possibilities of an evaluation might be.

So, if an evaluation does not make recommendations, what can it do to embed the legitimate voice of users and use into the evaluation? I refer here to an environment in which an external evaluation of a programme or intervention is commissioned. The key here is to have very early conversations about what the evaluation can do or is intended to be used for. Eleanor Chelimskey's (1997) distinctions between what she calls evaluation perspectives, but I would call uses is a good place to start. These are:

- Evaluation for accountability (e.g., measuring results or efficiency).
- Evaluation for development (e.g., providing evaluative help to strengthen institutions or projects).
- Evaluation for knowledge (e.g., obtaining a deeper understanding in some specific area or policy field).

Of course an evaluation might contain a combination of all these uses. The process of establishing the voice of the user at the outset of the evaluation and building into the design a real 'rehearsal' of the way in which an evaluation might be used, seems to me to be both neglected and important. Not to do so all too often results in missed opportunities for any use at all with evaluation reports languishing on dusty shelves overlooking meetings where decisions are being made.

It is useful to descend from the generality of Chelimskey's categories and imagine a real situation in which evaluators and users are sitting around a table and are

discussing what is going to happen to an evaluation report on an organizational change process. This means identifying a list of specific practices, for example:

- Tabling the report at a meeting to assess its implication.
- Deciding on what those implications might be and acting on them.
- Doing so in an agreed timeline.
- Undertaking staff development activities on the basis of the findings.
- Publicizing and disseminating more widely, etc.

The evaluator can help this process by presenting the evaluation report in such a way that the logic of the findings make it clear what the decision making options might be. A simplistic example to demonstrate this point might be the analysis of positive effects of a programme on the one hand and the cost on the other. Both these elements of the depiction should be faithfully recorded for the users in order for them to make a decision on what the balance should be. This is a useful act and the voice or needs of the users are present but it does not trespass into the realm of recommendation.

Using evaluations involve politics. While evaluators should not say what decisions might be made, they should be realistic and sympathetic to the world in which decisions *are* made. Users of evaluations have these considerations as central. Unpalatable findings might require presentation in a particular way, at a particular strategic moment to a particularly strategic audience in the first instance. Evaluators should be sympathetic to these realities and have little power to insist.

In presenting the voice of the commissioner, I have dwelt on the planning and design stages and the final stages of reporting of an evaluation. But, an obvious part of the obligation of an evaluation is to meet the expectations of users that the findings are strong enough to bear the weight of important decisions. In other words the depictions offered by an evaluation are valid and reliable. In my view normal social science standards pertain here. The way we collect evidence should be subject to those rules and conventions. This does not imply any social science paradigm, merely that the work complies with whatever standards are required for a given approach. What is important is that the claims that can be made on the basis of different approaches must be part of early conversations with users.

To conclude these observations on the user voice I would like to make two other points. There is an emergent culture in the 'developed' world that holds that the public seem reluctant to allow expert groups (teachers, health workers, judiciary, etc) to police themselves. The assumption used to be that their internalized professional values acted as a discipline on their practice and ensured the highest standards (in theory at least) and their pronouncements could be trusted. This aspect of the social contract has broken down. It seems that Governments, some would say the public at large, no longer trust these groups to discipline themselves (see Brown & Michael, 2002).

What has replaced it is a proliferation of evaluative mechanisms that are designed to control the quality and standards we used to allow the practitioners themselves to ensure. In this sense, evaluation is in danger of becoming a universal policeman.

This increases the burden on evaluators to take seriously their role in developing honourable practice (see Greene, 1999).

More optimistically, with this proliferation, we can see possibilities in the use of evaluation as part of social capital building. It has a role in the civic voice as it acts to bring states and organizations to account for the way they undertake their work and spend our money on our behalf. This goes some way in counterbalancing the policing role. Such evaluative practices are not the preserve of consultancies or university centres. Some of the most trenchant critiques of Government policy in the UK for example, are derived from the National Audit Office. Their most recent report on the UK Government's truancy initiative (See NAO, 2005), in which rates of unauthorized absences from school have remained the same after large amounts of money had been invested from the public purse is an example. What might be more important, however, is building social capital on a local, smaller scale, in which organizations become involved habitually in forms of reflective practice where there are mechanisms, checks and balances through which practitioners can make their voice heard. This issue leads to the next section in which evaluation's presence in social learning is considered.

Institutional and social learning

There is a range of activities that closely resemble evaluation and might legitimately be called a form of informal evaluation, remembering of course that informality does not preclude 'systematization'. It is to a discussion of evaluation's presence in these activities I now turn. The position adopted in this paper is that evaluation can take place within a set of social practices or within an organization as part of a cultural orientation rather than an over-evaluated or performance ridden raft of controlling measures and systems.

This is a reference to building an evaluative culture that has as a central tenet a series of reflective practices. This is not new in educational and professional development circles in which Schon (1991) and Michael Eraut (2000) for example, have identified the value of developing these processes in terms of organizational health, the adaptive capacity of organizations and in the development of professional identities that understand and are sympathetic to collaborative and creative responses to change. That tacit learning is continuous within social practice is axiomatic. The attribution of value and worth through judgements on what is professionally useful, rewarding or what works is part of social practice that can form the basis of such a reflective culture. In my view these processes of judgement are profoundly evaluative.

The knowledge resources we draw upon to be effective work people are complex (see Blackler, 1995). They are constantly evolving and can be subject to rapid change. How might we harness these resources for learning? I suggest this is where evaluation comes in. In a spate of recent projects (see Saunders *et al.*, 2005) in which the change process in institutions of HE in Europe was the subject of evaluation, a perspective was developed on the way in which evaluation might contribute to

institutional or social learning through the capture or depiction of creative approaches to the change process. The evaluation projects were, in effect, working within the tacit learning environment to which I refer above. This learning is the focus for these kinds of evaluations. It is based on the idea that we are working and learning in what social theory calls conditions of *chronic uncertainty*. I will return to this idea in the conclusion.

This vision of the use of evaluation to support planning and managing change is very different to the development of the influential movement variously called evidence based policy-making or evidence based change. This is a beguiling concept because it suggests a technocratic or rationalistic fix to the problem of change. We just find out what works, on the basis of evidence, and adopt that course of action. Others have pointed out that things don't quite work like that. Other research (see Wenger, 1999) suggests how context specific 'what works' tends to be and to reproduce the circumstances that brought about the positive changes often proves illusive. There are, however, technical problems with the role of evaluation in forming policy which have generated an interest in meta reviews or evaluations which attempt a synthesis of all the evaluations in a particular domain with the aim of distilling the key learning points. Again, we now know that this is not as straightforward as we might think in either syntheses of narrative evaluations or the integration of statistical studies (Pawson, 2002)

The presence of evaluation in social policy levels is identified at an earlier stage in this paper. The focus here is the potential, but as yet relatively ignored voice, of evaluations that can yield the resources for sense making through self evaluations. This territory is one in which there are unclear boundaries between institutional learning and evaluation, where evaluation can be embedded as a culture of reflection. This can be translated into some highly practical approaches from relatively informal embedded evaluation through review meetings and reflection to responsive approaches to quality enhancement and assurance as well as ideas such as the learning organization (see Burgoyne, 1999).

The presence of theory in evaluation

It is a cliché to write there is 'nothing so practical as a good theory'. I suggest that 'theory' can enter consideration at a very early moment in evaluation design. It does so in four ways that affect the extent to which evaluations contribute positively to social development. In other words, theories orientate an evaluation and determine the kinds of claims we might be able to make on its basis.

The starting point for me concerns the *theories of evaluation as inquiry* embedded in an evaluation either explicitly or implicitly. This is usually connected to a theory of social science. There are some obvious parallels, for example, between the use of random controlled trials or their variants in evaluation, for example, and a view of the social world that assumes there are social facts out there to be discovered and tabulated, i.e., a form of positivism. This is a method supported and practiced by Carol Fitzgibbon (1998) in the UK. This can be contrasted to an approach developed by

Saville Kushner (see Kushner, 2000) who suggests that constructing the stories or narratives of key participants and ‘personalizing’ evaluation as an important approach in the evaluation process. Whatever our predilection, we have to make some choices about how we connect to the social and technical world in order to identify what counts as evidence of what.

A second, rather more methodologically neutral (in the sense of the form evidence or data might take) concerns *theories of evaluation as a process*, i.e., the way evaluations should be carried out (mainly theories about how evaluations connect with elements or stakeholders in an evaluation process). Examples of this might be evaluation theory that espouses the advantages of a goal free approach, in which according to Scriven (1991), programme intentions are not relevant but the experience of a programme is. Michael Patton’s utilization focused approach (1997) to evaluation is another, in which key elements of a design explicitly express the interests and intentions of the evaluation commissioners and programme designers.

Thirdly, Carol Weiss (1997) and Connell and Kubisch (1996) discuss the way we should be looking for *the underlying theories about change that guide programme designs* and it is this that should form the focus of our designs. This can be understood as a ‘programme logic’. The idea of a bad change theory is an interesting one. There is a big debate in the UK at the moment on whether or not a pilot project approach embodies a good or bad theory of change. On what basis should a good example of practice, embedded in the special circumstances of a pilot, create wider changes? I have heard it rather pithily expressed that ‘the reality of a roll out from a pilot involves the application of fewer and fewer resources to more and more resistance’.

The fourth way in which theory has a presence in evaluation is through social theory of a more general kind. What is this resource? They are *theories providing explanatory frameworks for evaluation* that can structure our effort, suggest what kind of data is useful and, in my view, enhance the chance of the evaluation making a contribution to positive developments. The work of CSET for example is influenced by what can be termed ‘social practice’ theory (for an example see Wenger, 1999). It is an approach that emphasizes the situated activities and experiences of stakeholders that constitutes a programme in practice. This should form the centre of an evaluation and will yield the resources for judgement about value and worth. It also does justice to the diversity of experience and the voices of all those in the programme’s orbit.

I will illustrate this by reference to a metaphor embodying implementation theory called the ‘implementation staircase’. It has been recently used to design an evaluation for a project. The focus was a project called EQUEL (see <http://equel.net/>) as part of an EU (European Union) funded programme on e-learning. Its espoused purpose was to provide resources for a Centre of Excellence in E-learning. It involved a complex set of stakeholders all of which experienced the programme in different ways. The staircase metaphor is intended to capture the idea that the programme’s messages are understood and acted upon divergently by stakeholders as the programme’s message go down and up the ‘staircase’ of implementation. We understand the project implementation process as highly adaptive

and ‘practice’-based. This conceptual framework emphasizes the way in which policy messages are adapted and modified through the process of enactment. It is also important to understand the way in which policy messages are ‘transmitted’ through a system and are modified and adapted as they move from one group of participating stakeholders to another (see Saunders *et al.*, 2004)

Stakeholders occupying steps in the implementation staircase are both recipients and agents of policy and through this process the project message will undergo adaptation and be understood very differently according to the unique and situated experience of each stakeholding group.

Crucially, it is their interpretation of policy priorities, emphases and embodiments that are passed on to other stakeholders. There are two implications for evaluation here. First, policy should be depicted as multiple ‘policy in actions’ through the experiences of different stakeholders and second, policies and programmes are shifting and evolving entities depending on stakeholder experience of them. The metaphor of the ‘implementation staircase’ is used to capture this process of policy implementation (see Figure 2) in which messages go both down and up the staircase but are modified and adapted as they go. The evaluation design depicts the way the project was understood and enacted from the perspectives of the EU (the European Union funders), the project proposers and designers, the leaders of the various special interest groups (SIGs) and the SIG members. It suggests the importance of depicting the experience of the project from the points of view of all the main stakeholders within the process and positions the evaluation as the mechanism by which these captures might be made and fed back into the project process. Further, it

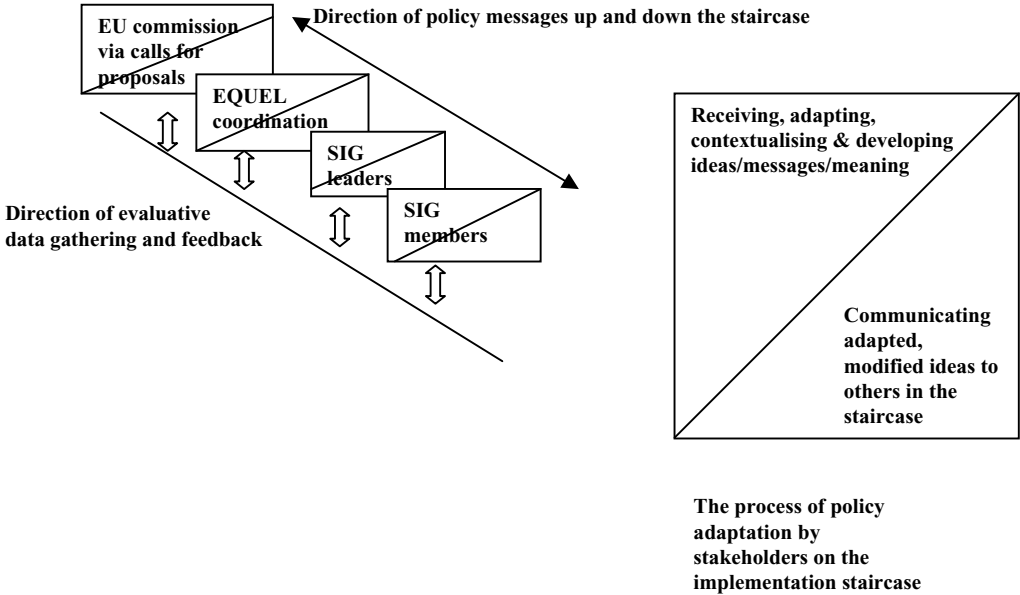


Figure 2. Implementation staircase

suggests these points of view may well differ significantly and it is the task of the evaluation to 'make explicit' these important differences. The evaluation was used by the participants as a way developing agreements and a core vision about what the project was for.

The international presence of evaluation

One of the clearest indicators of the way in which the voice of evaluation has become more mainstreamed in policy practice has been the growth over the last two or three years of the internationalization of evaluation. We have the formation in March 2003 of the IOCE (the International Organization for Cooperation in Evaluation). IOCE now has 63 regional and national societies with whom it has contacts. In the preceding year IDEAS (International Development Evaluation Association) was formed and there has been a rapid growth of regional forms of organization like the EES (the European Evaluation Society) that have joined more established societies like the AES (the Australasian Evaluation Society). At the same time, the UK is an example of this, the model of sub-regional organizations that express a responsive and more localized form of organization like the AES is gaining currency. Why has the international voice of evaluation begun to gain ground?

The starting point for this explanation connects to an earlier discussion on the positioning of external evaluation as a replacement for internal professional validation of the quality and value of provision. This tendency has a global hegemony. As Crawford *et al.* suggest (2004) in an analysis of methods for the evaluation of aid effectiveness, agencies are under increasing pressure to demonstrate that their focus and activities result in significantly positive effects. In many cases up to 10% of budgets are allocated to evaluation in programme funding. Government departments and international agencies attach evaluation criteria to initiatives in a way that that they simply did not before the 1980s. At the same time, agencies are beginning to adopt a more inclusive approach to evaluations and the voice of the 'recipients' through local as well as international consultants is louder. In some cases, in the US for example, aid projects are evaluated as a matter of law (Crawford, 2004, p. 175)

There is, therefore, an increased interest in accountability and learning in cross national funding activities involving evaluation partnerships that are also cross national at individual and institutional levels. A current evaluation of a joint DfID (the UK's Department for International Development) and UNRWA (United Nations Relief and Works Agency) funded intervention designed to support Palestinian schools 'as a focus for community development', is a case in point (see Saunders, 2001c). This evaluation involves a team from the UK and Jordan and includes a strong capacity building element (it continues as I write). In this kind of environment, donor and practitioner interests in the quality of and standards in evaluation have been pushed to centre stage. Evaluation has become an international preoccupation. How might the presence of evaluation be felt in this new internationalized context so that it can contribute effectively to development?

First, there is an imperative to collaborate across national boundaries around issues to do with protocols, procedures and ethics. This may be particularly important in what we might call 'low trust environments', put more bluntly, where evaluators and their evaluations might be put under inappropriate pressure to tinker with findings or conclusions or, more commonly, simply buried without a trace. While international cooperation might not in itself create power, it can be useful to have a set of guidance to refer to in times of trouble.

Secondly, there is an imperative to build capacity and to help train evaluators. This need does not simply refer to technical training in method or project management, it also involves designing new ways in which more experienced evaluators can work together with newer evaluators to induct them into communities of evaluation practice. This is the model developed in the Jordanian case (see above). The approach is built upon the notion of induction into a community of practice through engaging in a collaborative 'real-time' evaluation. There are important cultural and technical challenges in these collaborations including, in particular, contrasting understanding of the positioning and role of evaluation in development. In the Jordanian case for example, a culture of 'inspection' was dominant. This was gradually changed to include a more formative stance on the role evaluation could play in school development. The cooperative organizations in evaluation might proactively seek opportunities to undertake cross cultural evaluative work in which capacity building of this type can take place.

Thirdly, evaluators from developed countries should not be in the business of inadvertently promoting examples of western-oriented globalization through the monopolization of evaluation styles and expertise, which further entrench imbalances in the distribution of power and resources. To that end, international organizations, in evaluation or any other domain, should be careful to have some basic anti-globalization principles at their heart. It is worth drawing a distinction then, between globalization and internationalization. By globalization this paper refers to the process whereby national associations, communities and populations are brought together in economic, political and cultural relationships in which very uneven distributions of power and resources exist, often used to consolidate existing privilege. These relationships, more often than not, do little to improve the circumstances of the dispossessed and powerless. On the other hand, internationalization offers a prescription in which participation and inclusivity are actively sought where association provides a platform for a more 'conscious' approach to cross national activity.

The IOCE does aspire to these values. The vision of the IOCE is to promote cooperation between national and regional evaluation societies, associations or networks. Of course these sentiments will need to move beyond rhetoric to genuinely contribute to the development of enlightened evaluation worldwide. However, there are some important international trends that help to explain why the IOCE is a particularly apposite initiative at present and suggest it can play a useful role in the international development of evaluation (see Mertens & Russon, 2000). The development of the IOCE has as its core purpose to seek to legitimate and strengthen

evaluation societies so that they can better contribute to good governance and strengthen civil society. It aims to build evaluation capacity, develop evaluation principles and procedures, encourage the development of new societies and associations, seek resources for cooperative activity and undertake educational activities to increase public awareness of evaluation. It has the potential to be a forum for the exchange of useful and high quality methods, theories and effective practice in evaluation.

Defining an expansive presence for evaluation

Borrowing a term from the work of Yrjo Engestrom (see Engestrom, 2001), I am arguing for a place for evaluation in social policy and practice that expansively marshals 'knowledge', based on the experiences of stakeholders that can help chart a course through these changing times. This redefined expansive presence is not based on a myopic advocacy of one stakeholder interest over another but places a value on inclusivity for socio political reasons and, importantly, in the interests of sound evaluation designs. By these changing times, I mean conditions that are *chronically* uncertain. These conditions can produce, as Emile Durkheim (1952) observed over a century ago, periods of normlessness or anomie. Put more straightforwardly, periods in which the taken for granted knowledge we depend on to make sense of our world is shifting and, in some instances becoming destructively unstable.

New ways of learning and communicating for example, linked to the introduction of ICTs, can produce such instabilities as a *transition* is made across a *boundary* from one culture of practice to another. Practitioners and policy-makers can be supported by constructing *provisional stabilities* as they seek creative solutions to problems produced by change. What are provisional stabilities in this context? They are sense making knowledge produced by reflection on and the understanding of change, enabling choices or decisions for future action. The assertion is that we find it very hard indeed to proceed and plan for the future in conditions of rapid change. If we can capture and depict cases, examples, vignettes, even stories of the change processes within an organizational or policy setting, from a participants' point of view, this helps with sense making (Weick, 2001) and creates enough stability or knowledge of what is happening in order to make small onward steps. Evaluations in this sense can interpreted as 'boundary objects' (see Tuomi-Grohn & Engestrom, 2003) that help to make sense of the metaphorical 'space' that exists between one time and place and another.

In this way evaluation can provide expansive learning resources for participants and ways of making sense of the change process. It suggests a vision of formative evaluation that can provide for such reflections and act as a *bridging tool* for planning and innovation. This paper has argued that an inclusive approach to evaluation maximizes its opportunity to contribute positively to social and educational development. In developing this approach, we have a real chance of building an evaluative presence and devolving evaluative action to the frontline.

Notes on contributor

Murray Saunders is Professor of Evaluation in Education at the University of Lancaster, where he is Director of the Centre for the Study of Education and Training in the Department of Educational Research. He has worked as an evaluator in a variety of environments in both developed and undeveloped countries as well as those in transition. His research interests include theories of evaluation practice, higher education policy and change and the connections between education and work. Recent evaluations include the LTSN and the Subject Centre Network, the Quality Enhancement Framework for Scottish Universities and the evaluation of the learner's experience in the UK's Skills for Life policy.

References

- Blackler, F. (1995) Knowledge, knowledge work and organizations: an overview and interpretation, *Organization Studies*, 16(6), 1021–1045.
- Brown, N. & Michael, M. (2002) From authority to authenticity: the changing governance of biotechnology, *Health, Risk and Society*, 4(3), 259–272.
- Burgoyne, J. (1999) *Developing yourself, your career and your organization* (London, Lemos & Crane).
- Chelmskey, E. (1997) Thoughts for a new evaluation society, *Evaluation*, 3(1), 97–109.
- Connell, J. P. & Kubisch, A. C. (1996) *Applying a theories of change approach to the evaluation of comprehensive community initiatives* (New York, Aspen Institute).
- Crawford, P., Perryman, J. & Petocz, P. (2004) synthetic indices: a method for evaluating aid project effectiveness, *Evaluation: the International Journal of Theory, Research and Practice*, 10(2), 175–192.
- Durkheim, E. (1952) *Suicide: a study in sociology* (London, Routledge).
- Engestrom, Y. (2001) Expansive learning at work: toward an activity–theoretical reconceptualisation, *Journal of Education and Work*, 14(1), 133–156.
- Eraut, M. (2000) Non-formal learning and tacit knowledge in professional work, *British Journal of Educational Psychology*, 70(1), 113–136.
- Fetterman, D. M., Kaftarian, S. J. & Wandersman, A. (1996) *Empowerment evaluation: knowledge and tools for self-assessment and accountability* (Thousand Oaks, CA, Sage).
- Fitzgerald, F. (1972) *Fire in the lake* (Boston, Atlantic Monthly Press).
- Fitzgibbon, C. T. & Coe, R. (1998) School effectiveness research: criticisms and recommendations, *Oxford Review of Education*, 24(4), 421–433.
- Greene, J. (1999) The inequality of performance measurements in evaluation, *Evaluation: the International Journal of Theory, Research and Practice*, 5(2), 160–172.
- House, E. (1998) Institutional arrangements for evaluation, *Prospects*, 28(1), 109–116.
- Kushner, S. (2000) *Personalizing evaluation* (London, Sage).
- Mertens, D. & Russon, C. (2000) The inaugural assembly of the international organization for cooperation in evaluation, *The American Journal of Evaluation*, 21(2), 275–284.
- National Audit Office (2005) *Improving school attendance in England*. NAO, Report by the Comptroller and Auditor General.
- National Literacy Trust (2005) www.literacytrust.org.uk/Database/Primary/stratevaluation.html#Final.
- Patton, M. Q. (1996) *Utilization-focused evaluation* (Thousand Oaks, Sage).
- Patton, M. Q. (1997) Toward distinguishing empowerment evaluation and placing it in a larger context, *Evaluation Practice*, 18(2), 147–163.

- Patton, M. Q. (1998) Discovering process use, *Evaluation: the International Journal of Theory, Research and Practice*, 4(2), 225–233.
- Pawson, R. (2002) Evidence-based policy: in search of a method, *Evaluation: the International Journal of Theory, Research and Practice*, 8(2), 157–181.
- Saunders, M. (2001a) *An evaluation of 'children's rights are human rights': an interactive approach to children's rights education in Latin America* (London, British Council).
- Saunders, M. (2001b) *Focus on governance: British Council evaluation report on Chile* (Manchester, British Council).
- Saunders, M. (2001c) *Developing an evaluation approach in support to the education system of UNRWA* (Manchester, British Council, Manchester).
- Saunders, M., Knight, P., Goodyear, P., Fulton, O., Trowler, P., Jones, C. & Williams, S. (2002) *Evaluating the learning and teaching support network: from awareness to adaptation* (York, Higher Education Academy).
- Saunders, M., Charlier, B. & Bonamy, J. (2004) The evaluative research of complex projects in e-learning: the case of the 'EQUEL' (e-quality in e-learning) project, in: S. Banks (Ed.) *The proceedings of the networked learning conference 2004* (Lancaster, University of Lancaster), 733–745.
- Saunders, M., Knight, P., Trowler, P., Machell, J., Williams, S., Jones, C., Lent, N. & Allaway, D. (2004) *First annual report of the evaluation of the SHEFC quality enhancement framework* (Edinburgh, SHEFC).
- Saunders, M., Charlier, B. & Bonamy, J. (2005) Using evaluation to create 'provisional stabilities': bridging innovation in higher education change processes, *Evaluation: the International Journal of Theory, Research and Practice*, 11(1), 37–54.
- Schon, R. (1991) *The reflective practitioner: how professional think in action* (Aldershot, Ashgate).
- Scriven, M. (1991) *The evaluation thesaurus* (Thousand Oaks, Sage).
- Spector, R. H. (1985) *Advice and support: the early years of the United States army in Vietnam, 1941–60* (New York, The Free Press).
- Tuomi-Grohn, T. & Engstrom, Y. (2003) *Between school and work: new perspectives on transfer and boundary crossing* (Oxford, Pergamon).
- Weick, K. (2001) *Making sense of the organization* (London, Blackwell).
- Weiss, C. (1997) *Evaluation* (London, Prentice Hall).
- Wenger, E. (1999) *Communities of practice, learning, meaning, and identity* (Cambridge, Cambridge University Press).