

## Chapter 3

### Building capacity for social learning: what evaluation has to offer

*...in studying evaluation use, we began to observe that the processes involved in certain kinds of evaluations had an impact quite apart from the findings. In approaches to evaluation that involve participatory processes, those involved often experience changes in thought and behaviour as a result of learning that occurs during the evaluation process. Changes in program or organizational procedures and culture can also be manifestations of an evaluation's impacts. These observations about the 'process use' of evaluation led to a more direct focus on the potential of evaluation to contribute to organisational capacity development.*

Michael Quinn Patton 'The culture of evaluation' (Horton et al. 2003, p. v)

#### 3.1 Introduction

Chapter 2 enquired into the history, breadth and depth of social learning and found it to be increasingly regarded as a comprehensive concept, inclusive of a range of critical elements of complex environmental problem solving, i.e. learning and thinking elements, social and institutional elements, and elements of group participation and interaction (see Figure 2.1 & Table 2.5). I argue that the value of social learning to environmental management is the recognition of the relationship between these elements. It draws attention not only to practices of collaboration or learning but to the intersection of both collaboration and learning; or to the challenges not only of undertaking a systems thinking approach but embedding this within the real institutional constraints of decision-making structures.

A search in the literature also revealed a lack of consistency in theoretical development and praxis. In many ways this is not surprising. The practice of environmental management seems to be particularly bereft of active use or inquiry into relevant social theory. Many projects and programmes are initiated and run on the basis of professional experience, intuition and beliefs which remain unarticulated throughout the project, making it difficult to extract meaningful learning from contextual variability in comparisons across cases.

This chapter begins with the premise, then, that social learning offers a useful set of ideas to those purposed with the challenge of addressing complex environmental problems. However, some of these ideas are somewhat untested hypotheses and many lack tangible links between theory and practice. This suggests that the practice of social learning is in need of an active enquiry that develops both capacity and understanding. Such an active enquiry process must have the ability to reach into the wide-ranging aspects of the social learning concept, and return knowledge that is practical for immediate needs yet with the potential to be sufficiently profound to make contributions to the development of social learning as a concept. Added to this, it is ideologically consistent that the framework for ‘learning about social learning’ is one which is embedded in the actuality of the social learning process.

Authors such as Keen et al. (2005) have already identified reflection as a practice inherent to the success of social learning (see Figure 2.1). Others, e.g. Guijt (2008), have given their attention to monitoring as a way to improve the learning capacity of resource management dilemmas. Evaluation, particularly participatory and developmental (P & D) forms of evaluation, in essence combines both practices of reflection and monitoring. In my work as a researcher within CLEM, before beginning this thesis, I was already seeing the potential for evaluation to contribute to learning in environmental programmes. Evaluation practice, experience and theory has much to say on approaches that can influence the overall structure of a programme, provide guidance on what is going on – what is meant to be happening and what actually is – at the same time as enhance the learning capacity within the programme.

Furthermore, to build capacity for social learning requires having some influence on the institutional arrangements and social conditions of a given problem situation, particularly those which facilitate or constrain participation by diverse constituents, and which provide opportunities for experimentation and learning. Since programmes aimed at intervening in environmental problems (whether they are discrete projects or long-term endeavours) are both a manifestation of existing social norms and theories of action and an attempt to make changes in the social conditions of problem situations, they are a means by which these particular social learning conditions are shaped. Hence, evaluation – where it is so constructed to provide developmental support to programmes, and enquiry into institutional and social factors that

influence programme activity – can be a means of influencing these conditions for social learning.

In this chapter I look into the literature on evaluation, highlighting particular branches of evaluation theory and practice that hold most promise for building capacity for social learning in environmental management programmes. Firstly, I outline what is meant by building capacity for social learning. Secondly, I explain why the field of evaluation is relevant to capacity building for social learning and what the most recent developments in evaluation theory and practice have to offer. Thirdly, I postulate a relationship between evaluation and social learning which will be used in the examination of the case stories in the subsequent chapters.

### **3.2 Building capacity for social learning – what does this mean?**

Capacity building is a term which has had currency for sometime among academics and practitioners in community development, and increasingly it is appearing in discussions around behaviour change and environmental management. At its most basic, capacity building refers to activities that improve the ability of either an individual or organisation to achieve its goals (Linnell 2003). However, Allen (2007) points out that capacity building is only really meaningful when it is discussed in reference to what you intend to build capacity in; and the range of possible contexts is vast. Capacity building can take place across organisations, within communities, or in whole geographic areas. It can involve individuals and groups of individuals, organisations, groups of organisations within the same field or sector, and organisations and actors from different fields and sectors (Linnell 2003).

With such a broad range of potential contexts in which to build capacity, the focus of capacity building is similarly extensive, ranging from infrastructural matters such as core funding, resourcing, and providing expertise support, right through to technical training and facilitated organisational development. Who gets involved in capacity building is correspondingly wide-ranging. Primarily it is undertaken by external agents, such as government bodies, foundations and professional associations (Cigler 2001), but it can also be carried out by management

consultants, grant makers, researchers and academic centres, or specific intermediaries and umbrella organisations (Linnell 2003).

Historic and ongoing critiques of capacity building initiatives include concerns that they have traditionally focused on technical matters rather than social process skills such as the ability to problem solve, work collectively, manage conflict, or deal constructively with matters of power and influence. Furthermore there is a growing wariness of top-down capacity building approaches (Ford et al. 2001; Andrew & Robottom 2005). This can refer to the practice of experts coming into a situation to impart knowledge without cognizance of existing skills or interest in building self-learning capabilities. It can also refer to a critical matter at the heart of all capacity-building initiatives – ‘who determines the agenda?’ For instance, in a story of tensions between Australian landowners and government over the conceptualisation of sustainability, during the 1960s, 70s and 80s, Andrew and Robottom (2005) poignantly illustrate what can happen when tacit assumptions about what it is important to build capacity in go unexamined. In this example, despite tapping into the best expert advice, farmers faced increasing problems of drought and soil erosion, as the capacity-building initiatives at the time stemmed from a government-led agenda not related to these concerns but rather directed towards increasing economic productivity (ibid., p. 66).

Contemporary writers and practitioners involved in implementing or theorising on capacity building for complex social systems and situations, such as community health, and environmental management, stress that in these contexts technocentric linear-information-transfer models of capacity building are inadequate (Cigler 2001; Andrew & Robottom 2005; Allen 2007). As Allen (2007) notes, the central concerns of environmental management and social learning *to manage change, to resolve conflict, to manage institutional pluralism, to enhance coordination, to foster communication, and to ensure that data and information are shared, require a broad and holistic view of capacity development.* Further Cigler (2001), working in the field of multi-party networks and collaborations, points out, that while it is important to acknowledge that successful communities of the future will chart their own course, based on their particular characteristics — they appear to need help to do so. In which case *innovative types of capacity building that hone collaborative skills play important roles in*

*preparing communities and their organisations for the changes associated with complex partnerships* (ibid., p. 83).

What is emerging from the critique of technocentric linear-information-transfer models of capacity building is an interest in linking capacity building to the growing understanding around participatory and empowerment-based learning, emerging in circles of community development, and community-based environmental and health programmes (e.g. Horton et al. 2003). However, even capacity building initiatives conceived within this more socially aware paradigm are subject to further concerns over matters such as the costs of capacity building (which can make it prohibitive for many); difficulties with prioritising capacity building against other strategic objectives, and an underlying uncertainty captured by the question ‘what exactly is the role of the professional capacity builder? In recent decades the evaluation community, has been grappling with parallel challenges of limited resources and low priorities for their work. It has also faced comparable changes in their role – tracking shifts from the evaluator as external analyst and critic, to recognising that the evaluator has a unique opportunity to promote learning in programmes, and empower programme participants to make changes for themselves. This new character of evaluation, where it increasingly intersects with capacity building, is examined in the following sections.

### **3.3 Overview of the development of evaluation theory and practice**

Evaluation owes its origins to the perceived need in the 1960s to find ways to track the progress of government policies, interventions and programmes. However, over the 40-year history of evaluation as a recognisably independent field, concern has grown within the evaluation community regarding the use and value of their work. Notions of what it means to have evaluations ‘used and of value’ have also shifted, and there has been augmented expectation that evaluation yield not just an analysis of ‘what happened’ but results in increased learning at individual, project and institutional levels. This interest in the learning impact of evaluation has been a driving force shaping the divergent trajectories of evaluation theory. However, this has not been the only influence on emergent theories in evaluation. Evaluation has been subject to what amounts to methodological schisms as a result of the widely different contexts in which it has begun to be applied and as a consequence of some fundamental ontological differences.

Mapping a path through this is a somewhat daunting task and my version of events has chosen to focus on the impact on evaluation theory and practice caused by some specific shifts in viewpoint that link well with social learning<sup>1</sup>. Three shifts of viewpoint of particular interest are:

- Expansion of the core drivers of evaluation from client concern with accountability and information generation to evaluator interest in learning and organisational change
- Expansion of focus from producing evaluation **outcomes** that are valued and used to developing evaluation **processes** that are valued and used
- Increased cognizance of the power issues and potential for learning and development associated with evaluation knowledge.

These three trends of thinking have had transformative impacts on the field of evaluation resulting in innovative development of evaluation approaches.

### 3.3.1 Definitions of evaluation

Fundamentally authors agree that the concept of evaluation refers to a systematic assessment of a situation at a given point in time, past, present or future (Twomlow & Lilja 2004). Beyond this there is almost immediately a departure from consensus. The literature is rich with divergence on purpose, theoretical framework, underpinning ontology and, naturally, the implications this has for method. For instance, a seemingly relatively straightforward statement by a New Zealand evaluation researcher about the nature of evaluation implies that judgment is the central function:

*Evaluation is the process by which we examine, assess and make judgments about the relative or absolute value of an action, a process, a practice, or an investment.*  
(Saville-Smith 2003, p. 16)

However, authors Guba and Lincoln (1989a) describe four generations of evaluation practice that have emerged over the years and identify a judgment function in evaluation as a

---

<sup>1</sup> It is important to note that while these are shifts of viewpoint, they result in an expansion rather than change in the evaluation field. This is because as evaluation has developed it has embraced a wider range of drivers, focus and methods rather than replaced historical ones with newer versions.

characteristic of 'third generation' evaluation. This has been preceded by frameworks based on accounting and description (still widely used) and superseded by frameworks that respond to the learning and information needs of multiple stakeholders. In other words the idea that evaluation is about judgement has only come about after many years of working with evaluation methods that have concentrated on describing, enumerating and measuring various aspects of public policy interventions. Further, many modern evaluation practitioners and theorists have started to work with new ideas about the purpose of evaluation, i.e. that it is about enabling a programme and its participants to learn, adapt and respond to the needs of the situation.

Even categorisation of evaluation is not universally agreed upon. While the earlier mentioned classifications of first-, second-, third- and fourth-generation evaluation described by Guba and Lincoln (ibid.) would be widely recognised amongst the evaluation community, there are many other ways in which authors have chosen to make distinctions. For instance Baehler (2003, p. 31) divides evaluation into process type and impact/outcome type, and Duignan (2003, p. 77) makes a distinction between policy and programme evaluation. In this age of pluralism, a multitude of others have introduced entirely new frameworks which they become predominantly associated with. Widely known examples are naturalistic/responsive (also known as fourth generation) evaluation (Guba & Lincoln 1989a); theory-based evaluation (Weiss 1995; Stame 2004); realistic evaluation (Pawson & Tilley 1997); participatory evaluation (Brunner & Guzman 1989) and empowerment evaluation (Fetterman 1996). Added to these are numerous less widely discussed contributions to evaluation theory and practice which also adopt new branding terminology to distinguish their ideas. Examples include the evaluation voices method (O'Sullivan & O'Sullivan 1998), partnership evaluation (Oliver et al. 2003), and evaluative enquiry (Preskill & Torres 1999). These varying approaches to evaluation are not sequential evolutionary developments, and frequently share as many characteristics as they have differences.

*Summative, formative, cost-free, goal-free, functional, tailored, comprehensive, theory-driven, stakeholder-based, naturalistic, utilisation-focused, pre-ordinate, responsive and meta* are but a small set of the terminology that could be attributed to evaluation (Pawson & Tilley 1997, p. 1). While many authors describe evaluation as a comparatively young discipline (ibid.; Saville-

Smith 2003) it is widely acknowledged to have grown exponentially in the face of a global trend towards decentralized bureaucracy and control through surveillance (Pawson & Tilley 1997, p. 1). Further evaluation has developed in multiple contexts, e.g. public health, education, community development and organisational change. It would be easier now to list areas of governance, funding and research which do not have a branch of evaluation theory and methodology directly contributing to it.

Duignan (2003) points out that evaluation undertaken in many different disciplines can be conceptualised in a number of different ways and the variety of terms can lead to problems in discussing, commissioning, undertaking, and reporting evaluations. He goes on to identify four conceptual levels for divergence in evaluation terminology: the evaluation approach, purpose, methods and designs (see Table 3.1).

**Table 3.1 Conceptual levels for evaluation terminology** (from Duignan 2003, pp. 78–79)

Evaluation approach	Overall way of conceptualising evaluation including philosophical and value orientation to the task	e.g. Kaupapa Māori <sup>2</sup> , goal-free, utilisation-focused empowerment)
Evaluation purpose	Sometimes called ‘types’, this refers to the intended use of the evaluation	e.g. formative, process, impact/outcome, summative
Evaluation method	Ways of carrying out research	e.g. surveys, focus groups
Evaluation design	Way in which the methods of an evaluation are used to answer evaluation questions for a particular purpose under the overall framework of an evaluation approach	The design may use a number of different types and purposes within an overall approach.

Adding to the confusion is the disjunction between what is theorised about evaluation and what is actualised through evaluation practice. Chen and Rossi (1989, p. 299), note that the paradigmatic shifts that have occurred in evaluation largely operate at what they describe as the *high culture level*, rather than at the level of the everyday operative evaluator. In fairness, the blame for this is as likely to be constraints imposed by the limited expectations of the commissioner and funder of the evaluation, as the failure of the professional community to take

---

<sup>2</sup> Using Kaupapa Māori as a basis to evaluation implies that the specific philosophical and cultural dimensions of working with Māori communities are at the forefront of the evaluation approach.

up the theoretical advances on offer. However, the pragmatics of undertaking an evaluation in the end can make it likely that the best an evaluation theorist can hope for is that evaluators may use key features of their frameworks as guidance to their practice. As Davidson (2003, p. 106) points out: *real evaluation is a messy process characterised by numerous false starts, frustrations, and a lingering worry that you've chosen the wrong method in the first place.*

In summary, writing as far back as 1989, Chen and Rossi make a neat statement that has proved to be prophetic: *evaluation has never been dominated by a single paradigm, nor is it likely to be in the foreseeable future* (p. 299). Furthermore there are certainly those who regard epistemological eclecticism as a strength not a weakness and consider the secret to success is to maintain a vigilant reflexive approach to practice, staying aware of and questioning assumptions (Davidson 2003).

### **3.3.2 Critiques and new directions for evaluation**

In 1980 Cronbach et al. observed that *an evaluation ought to inform and improve the operations of the social system* (p. 152). In his work on reform of programme evaluation Cronbach and his colleagues discuss expanding the role of evaluator beyond that of an investigator who brings in technical skills to a circumscribed problem, communicates findings in a report and departs. Their alternative vision is one where the evaluator is actively engaged in the political events of the situation, working as a *multi-partisan who serves the general interest* (ibid., p. 152). Cronbach's evaluator, in his terms, is an *educator*. Through *holding the mirror up to events* (ibid., p. 153) the evaluator gives decision-makers the chance to make decisions with greater awareness of the complexity of phenomena, the underlying assumptions, and the long-term possibilities as well as short-run advantages of situations. Operating in this education-based approach Cronbach indicates that what evaluation can offer community development programmes is lessons on how to choose what variables to observe to improve their function and delivery (ibid., p. 169). Furthermore the evaluator, by moving amongst the many programme constituents, cross-pollinates ideas from a range of stakeholders, clarifies the multiple objectives and can help to redefine the problem context for the programme (ibid., p. 171). The success of the evaluation is then to be judged as any other educational effort *by what it does to develop the learner's potential* (ibid., p. 160). To understand that in 1980 this was still

deemed a radical departure from conventional evaluation practice it is necessary to look at the trajectory of changes to evaluation theory and practice over the preceding decades.

In the late 1980s Guba and Lincoln (1989a) proposed a radical new construction for evaluation which they regarded as the fourth in a series of generations, each characterised by shifts in focus and changes to the role of the evaluator (*ibid.*). Evaluation can generally be said to have emerged out of drives towards greater rationality in government social policy and a perceived necessity for improved information for decision-making. What was then common to early and long-time dominant forms of evaluation was a focus on measurement. Hence Guba and Lincoln (1989a, p. 26) refer to first-generation evaluation as the *measurement generation*. Here measurement and evaluation are interchangeable terms. Variables are predetermined and the role of the evaluator is a technical one, providing a range of instruments (largely quantitative) to gauge them.

Spawned from perceived deficiencies in the usefulness of information generated solely around the measurement of predetermined variables, second-generation evaluation developed a focus on *objectives*. It is characterised by the description of strengths and weaknesses in relation to predetermined programme objectives. Measurement becomes one of the tools of evaluation and is no longer synonymous with it. The role of the evaluator in second-generation evaluation is as a *describer* (*ibid.*, p. 27). However, what soon becomes apparent is an absence in all this description of any capacity for judgment. This is further exacerbated by reliance on pre-formed objectives, the validity of which only becomes apparent after the project is completed. Therefore the shift in focus in third-generation evaluation is towards *decisions* (*ibid.*, p. 31). The development of a judgment function in evaluation necessitates a judge – a role which evaluators have reluctantly taken up, and found to be, as expected, rife with problems as it plunges them directly into the political arena and compromises the hitherto much prized objective independence (*ibid.*).

Guba and Lincoln identify a set of fundamental and interrelated problems with all three evaluation generations. These are: (i) a tendency toward managerialism, (ii) failure to accommodate value-pluralism, and (iii) over commitment to the scientific paradigm of inquiry (*ibid.*, pp. 32–37). Although these are couched here in Guba and Lincoln's terms, observations

on each of these areas of critique have been made by many other authors in the field (e.g. Pawson & Tilley 1997; Oliver et al. 2003).

1. **Tendency toward managerialism.** This refers to inherent problems in the relationship between the evaluator, the manager of the project under review, or the commissioner of the evaluation (often the project funder). Evaluations are often set up as independent, expert-driven, processes reliant on an established hierarchy and set of roles. For example, the community are seen in terms of *problems*, the policy analyst devises *solutions*, the service provider *implements* these, and the evaluator *observes, reviews and makes judgment* (Oliver et al. 2003). Such an evaluation model effectively avoids any questions about the practices and qualities of the manager and funder of work. Since these are the same people who generally have control over the dissemination of any results, the consequences are disempowerment, and disenfranchisement for both the evaluator and other project participants.
  
2. **Failure to accommodate value-pluralism.** Guba and Lincoln (1989a, p. 34) observe that evaluations are commonly regarded as scientific and therefore value free. However, in their opinion, in practice, this is far from true. A simple example of this is the tendency for the project manager or funder to set the agenda of the evaluation. In such cases, programme managers, affected by desire to make a ‘good showing’, are consciously or unconsciously selective in the variables they dictate for evaluation.
  
3. **Over commitment to the scientific paradigm of inquiry.** Although not alone, Guba and Lincoln are among the foremost in proposing a shift from positivist-based evaluation to a constructivist approach. In their view this positivist-based drive for generalisable results has led to *context-stripping*, i.e. ignoring the environment of the subjects of the evaluation. It also results in an overdependence on formal quantitative measurement, and an abdication of responsibility on behalf of the evaluator, because they are simply revealing ‘the truth’.

Stame (2004, p. 59), in her review of theory-based evaluation, suggests that the positivist tradition, in its reluctance to engage with the issue of values, has limited the ability of evaluators to contribute to discussions on the theoretical implications of programmes. Instead evaluators have concentrated on developing methods that test the internal and external validity

of a programme (i.e. did it do what it set out to do and can you generalise about the results?) and steered clear of challenging the theoretical rationality of the programme.

What Guba and Lincoln offer as a remedy to previous generations of evaluation practice is fourth-generation evaluation. This is based on constructivist understandings and hence represents a radical departure from positivist-based-evaluation formulas. Here programmes are recognised as social events and, by necessity, interpretations of these events must be negotiated with stakeholders who are the primary focus of this form of evaluation. Thus *the engine of the method is...an exchange of meaning between the researcher and all program participants* (Pawson & Tilley 1997, p. 18).

Pawson and Tilley (1997), in their own synopsis of methodological change in evaluation research, echo these critiques and offer a different, but not contradictory version of events. Writing some time after Guba and Lincoln's fourth-generation-evaluation work had appeared, they include this among three significant developmental phases of evaluation theory and practice. Terming the first phase of evaluation as experimental they describe it as based on a theory of causal explanation (i.e. the classic positivist science paradigm of control groups and measurable interventions). Experimental-based evaluations were designed to assist rational choices in policy around the best options, but their tendency was to end up with a complex of controls to try to ensure the validity of causal claims.

This is followed by the pragmatic phase, which is more grounded in the realities of policy making. Here the flow of knowledge is opposite to that of experimental-based evaluation, i.e. it starts with an understanding of the needs of policymaking and ends with knowledge that is considered valid if it is pragmatically acceptable within the set policy framework. The objective is to enlighten rather than provide definitive options, and hence the focus of the evaluation is clearly on the policymaking community. The third phase is Guba and Lincoln's aforementioned fourth-generation (naturalistic) evaluation.

Pawson and Tilley (ibid.) see problems with all three approaches. In particular they note (in agreement with Guba and Lincoln) that the first strips away context by being experimental and control based, but the last, as an opposite extreme (Guba and Lincoln's own suggestion), is of

compromised value because it places context so highly it is not possible to extrapolate and generalise. What Pawson and Tilley (1997) offer as an alternative is what they term realistic evaluation, which, in their view, draws both context and theory together to offer meaningful interpretation of events.

From all this reformist activity it may be easy to assume that the old ways of evaluation have largely been abandoned. This is far from the truth. Authors writing about new approaches to evaluation recognise them as departures from the norm, and even in some cases representing currently marginal viewpoints within the evaluation community. As Twomlow and Lilja (2004, p. 1) point out *[t]he way most evaluations are commissioned and conducted still aims at making definitive judgements about project worth rather than providing learning opportunities*. The one driver that evaluators share across the board, whether they are undertaking traditional cost-benefit analysis or radical soft-systems based participatory evaluation, is the desire to see their evaluations put to use. The cry of the evaluator, since the first accounting study was undertaken, was surely one of dismay at the lack of impact their patiently gathered information had in the critical decision-making it was designed to assist.

Being more influential with their work was certainly a driver behind the pragmatic evaluation trend described by Pawson and Tilley (1997) where evaluators seek greater engagement with the policy world. For others this has led to increased efforts to improve communication between themselves and those influencing decision-making processes, including an increased interest in understanding how organisations use information and learn. For instance, in his paper on evaluation and organisational learning, Rist (1997, p. 19) reflects on the question *why is it that organizations appear to have more receptivity to certain types of information and to the manner in which it is packaged than they do to other types?* and posits that the evaluation community needs to consider how to restructure their work to respond to the dynamics of organisational learning. Similarly Preskill and Torres (1999, p. 94) found they were driven to develop a more meaningful role for evaluation in organisational development after they *increasingly encountered instances where our work went unused and realized what little impact traditional evaluation practice was having on organisational change efforts*.

In their discussion of evaluation use Forss et al. (1994) make a helpful distinction between types of use that are instrumental and those that are conceptual. Instrumental use is where the results of an evaluation are used to make decisions and change projects. Conceptual use, in contrast, implies that the people involved are affected in how they think about an issue<sup>3</sup>. My own experience with evaluation suggest that an interest in enhancing instrumental use tends to focus evaluations on ‘outcomes’, whereas an interest in conceptual use leads developers of evaluation approaches to consider ‘process’ and its relationship to learning to a greater degree.

Fundamentally then, where evaluation theory and practice have developed in different directions it is not because there are not some widely shared views about the limitations of traditional forms of evaluation, and the need to improve the use of evaluation findings, but because evaluation theorists and practitioners have differed on what they consider to be the most important issue to address, or most promising way to go about it. While there are those who consider that the solution to ensuring evaluations are used lies in finding ‘the right communication approach’, or perhaps the right person to talk to, of more interest to the quest in this thesis (to find evaluation approaches that works well with the demands of building capacity for social learning) are those evaluation theorists who respond to a shift in view of the role of the evaluator from independent technical advisor to engaged facilitator of learning and change. These theorists often have in common an interest in the move from externalised accountability-based evaluation to internalised improvement-based evaluation (Torres 1994, p. 333). There are a number to choose from and the next part of this chapter will concentrate on reviewing their work for its applicability to the question of developing social learning capacity in complex problem solving.

**In summary**, Table 3.2 outlines fundamental shifts and drivers that have shaped evaluation theory and practice. They mark the altered perception of evaluation from an instrument of **accountability**, to understanding its potential to create **improvement** in programme functioning, to ultimately, regarding evaluation as an intervention that can fundamentally affect the capacity of programmes to deliver **change and development** in their target contexts.

---

<sup>3</sup> Weiss (1977) refers to this conceptual use of evaluation as *enlightenment* and Berk and Ross (1977) refer to it as *demystification* (cited in Forss et al., 1994, p. 576).

Overarching shifts in purpose for evaluation coincide with changed emphasis in evaluation practice: from generating information, or supporting judgements about programme efficacy, to influencing programme learning. This has been accompanied by altered focus, from ensuring the client or funder needs are met by communicating the findings of the evaluation, to strengthening relationships between the evaluation and the policy context for the work, and most latterly paying increased attention to the needs of stakeholders who implement the programme or are the intended beneficiaries. Methodological changes include shifting from a positivist epistemology to a constructivist basis for evaluation with all this implies; and recognition of the need to consider both the instrumental and conceptual use of the evaluation. This in turn has implications for the role of the evaluator as a technical expert or a skilled facilitator of learning within the programme.

**Table 3.2 Changes and trends in evaluation (1960s to present day)**

<b>Accountability</b>	<b>Improvement</b>	<b>Change &amp; Development</b>
Information (description, measurement)	Judgement (comparisons, matching objectives to outcomes)	Learning (facilitation, frameworks)
Focus on ‘outputs’ that are valued and used	Focus on ‘outcomes’ that are valued and used	Focus on ‘processes’ that are valued and used <i>e.g. orders of outcome evaluation</i>
Focus on client/funder needs (improve communication)	Focus on policy needs (improve evaluator/policy interface)	Focus on stakeholder/participant needs (improve stakeholder–evaluator interface) <i>e.g. participatory evaluation</i>
Emphasis on instrumental use		Emphasis on conceptual use <i>e.g. theory-based evaluation</i>
Evaluator as technical expert		Evaluator as facilitator of learning <i>e.g. empowerment evaluation</i>
Positivist epistemology		Constructivist epistemology <i>e.g. naturalistic evaluation</i>
Works within implied theory and logic of programme		Develops theory and logic of programme <i>e.g. theory-based evaluation</i>

In Table 3.2 the far-right column highlights recent developments in evaluation theory and practice of most interest to building the capacity of social learning in environmental management programmes. Included here are examples of evaluation approaches that have responded to these drivers and shifts in focus. For example participatory evaluation has much to say about the involvement of stakeholders in evaluation. Empowerment evaluation extends this further into making these stakeholders the masters of their own evaluation process.

In the final row of the table I add one further path of change in evaluation – a movement from accepting the implied rationale and theory of the programme as a basis for evaluation, to recognising that this at times may be the root of programme failure and that evaluation has a role to play in introducing new theory and improving the logic of the programme. This capacity to introduce theory and expose tacit assumptions that drive programmes is an important potential avenue for introducing concepts of social learning into environmental management initiatives. A cluster of approaches termed ‘theory based evaluation’ have developed in response to this perceived opportunity for evaluation.

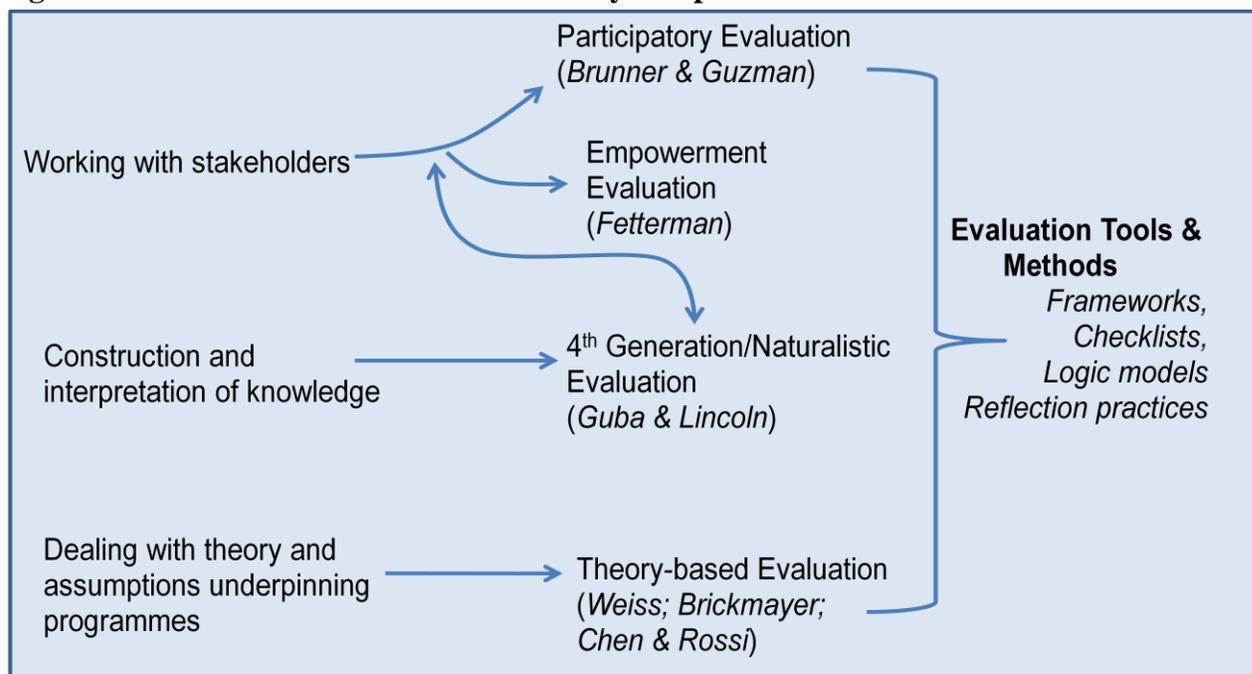
### **3.4 Evaluation’s contribution to capacity building for social learning**

Learning and change are clearly not always the foremost directives of evaluation. Many, if not most, evaluations undertaken have a primary function to gather information for decision-making, paying varying degrees of attention to the links between that information gathering and the decision-making processes. Indeed there are many purposes behind undertaking evaluation which may even act as obstacles to learning, such as legitimisation, camouflage, or to develop ammunition in a struggle for power (Forss et al. 1994). However, the previous section reviewed developments in evaluation theory and practice that show promise for a useful intersection with social learning. These developments have emerged in response to multiple drivers but essentially in line with the three shifts in viewpoint identified earlier: that is (i) changed drivers, i.e. from client concern with accountability and information generation vs evaluator interest in learning and organizational change; (ii) changed focus, i.e. from evaluation **outputs**, and **outcomes** to evaluation **processes**; and (iii) increased cognizance of power issues associated with evaluation knowledge.

While no one branch of evaluation has addressed all three with equal emphasis, these shifts of view have been formative in several evaluation approaches. Figure 3.1 outlines the four branches of evaluation theory and practice that will be explored in the remainder of this chapter. The figure also lists names of one or more theorists whose work has been seminal in the development of each evaluation approach: (i) changes in the way stakeholders are regarded and involved in evaluation practice forms the basis of participatory evaluation; (ii) even further commentary on the implications of this for the role of the evaluator is offered through empowerment evaluation; (iii) although it has much to say about stakeholder engagement fourth-generation evaluation has been included here principally for what its commentary on how evaluation approaches can contribute to the development of knowledge; (iv) theory-based evaluation has emerged in response to a perceived need to find ways to improve the theoretical basis and inherent logic of programmes and interventions.

Finally each of these evaluation approaches is, in practice, implemented through tools, techniques and methods that are themselves worth scrutinising for how they can contribute to the development of the social learning potential of environmental management programmes.

**Figure 3.1 Four branches of evaluation theory and practice.**



### 3.4.1 Increasing stakeholder participation in evaluation

The first branch of evaluation approach discussed here is that which has developed primarily in response to the perceived need for increased stakeholder participation.

#### Participatory evaluation

Brunner and Guzman (1989), in their paper on participatory evaluation – *a tool to assess projects and empower people*, cite a familiar cast of grievances against traditional project evaluation for its *insensitivity to their true achievements and real problems* (ibid., p. 9). Their criticisms include that local projects feel victimised by evaluators, evaluations reflect the worldview and priorities of sponsoring agencies, and main actors are denied meaningful input while the focus is on pre-coded questionnaires, cost–benefit analysis, observation sheets, and quantitative data. These can readily be seen as the pragmatic and observable consequences of the problems with conventional evaluation commented on by Guba and Lincoln (1989), namely, the tendency to managerialism, failure to accommodate value pluralism, and overcommitment to the positivist research paradigm.

Lack of reflection on who benefits, and who **should** benefit out of evaluation is still a prevalent issue today. Davidson (2003, p. 102) states that, at the most basic level, there are three considerations which drive the choice of a particular research approach or method: (i) What do you want to know? (ii) From whom do you want to know it? (iii) How many resources do you have to find it out with? In my view, what is clearly missing from this list is the question ‘Who needs to know it?’ By placing the evaluator at the centre of the enquiry there is no opportunity to reflect on who the evaluation is serving, i.e. is it the evaluation commissioner, the project managers, or the participants? The evaluator becomes a knowledge broker, gathering and disseminating without need to be cognisant of the power dynamics inherent in such an arrangement.

With an unashamed social reform agenda in mind Brunner and Guzman’s (1989) ‘participatory evaluation’ aims to put the ‘who’ of development projects firmly in the picture when it comes to their evaluation. Participatory evaluation shares theoretical foundations, ideological convictions, and methodological principles with movements of popular education and participatory research in the Third World. Furthermore Brunner and Guzman express the

intention of empowering people through participatory evaluation to *join the struggle for a just and egalitarian society*. They assert that evaluation should be *permanent, participatory and educational* (ibid., p. 10).

The aim of such evaluation is to help beneficiary groups and local facilitators adjust strategies. Hence the ingredients of their participatory approach to evaluation include, foremost, a shift in the roles of evaluator and evaluatee. In participatory evaluation the evaluators are the principal actors of a development project (the groups conventionally called the target population or beneficiaries), while professional evaluators act as methodological consultants rather than decision-makers. In the evaluation the groups and the evaluators decide collectively what should be evaluated, how the evaluation should be carried out, and what should be done with the results. Evaluation is used to control the progress of the project, explain its problems, and establish consensus on what to do next. Preliminary results are shared with all people who have a stake in the project; reactions to or interpretations of the results are recorded, organised by the evaluation team, and turned over to the groups responsible for making the decisions for future action. Such evaluation is formative in nature and its success is based on the constructive action that it generates and the improvements made in the programme. There is also clearly an intention to improve collective reflection among project participants. This is designed to help clarify divergent values and norms that can prove influential in a project's progress, and even beyond that, to produce action-oriented knowledge that is based on shared norms and a common world view.

Brunner and Guzman readily acknowledge the challenges in undertaking participatory evaluation. Principal among these is its radical approach toward empowerment. They observe that participatory evaluation can only be successfully implemented where the institution that promotes it truly wants to emancipate the dominated groups and when the groups are prepared to assume responsibility for it (ibid., p. 16). Furthermore they note that proponents of more traditional methods of formative evaluation are often concerned that standards of scientific knowledge generation are not met and that participatory evaluation produces only subjective knowledge which is not suitable to explain change or be used for policy decisions.

A further challenge to this praxis-oriented approach in application is that the perception of reality among community groups and facilitator teams is conditioned by experience and their culture. Their appreciation of a situation, therefore, may be biased by a collective misperception, which can result in misguided action. Brunner and Guzman acknowledge this risk and state that participatory evaluation tries to diminish the danger of this through promoting a permanent attitude of critical reflection and by frequent evaluation exercises (ibid.). Importantly, participatory evaluation has profound implications for the role of the evaluator, and the skill set they require. It is no longer sufficient for evaluators to be versed in methods of assessment; they must have the capacity to impart these techniques to others, and to facilitate their learning.

There is now a vigorous community of practitioners of participatory evaluation, and although this approach emerged primarily within the development community context it has extended into numerous other arenas, including health and education. It has also spawned variations upon its main themes, designed for specific issues and environments. In New Zealand, this includes the ‘partnership evaluation’ approach described by Oliver et al. (2003). Developed within the context of working with Māori and Pacific Island communities, it shares the same principles as participatory evaluation but with added focus on establishing partnership between evaluators, programme funders and service providers.

### **Empowerment evaluation**

Almost indistinguishable at first glance from participatory evaluation (but without quite the same degree of emancipatory fervour) is Fetterman’s ‘empowerment evaluation’ (1994, 1996). Fetterman describes empowerment evaluation as *the use of evaluation to help others help themselves* (1994, p. 305). It is designed to foster self-determination rather than dependency, focuses on improvement rather than measuring outcome, is collaborative, and requires both qualitative and quantitative methodologies.

Self assessment is a critical component of empowerment evaluation: [It is] *pervasive...built into every part of a program, even to the point of reflection on how its own meetings are conducted and feeding that input into future practice* (ibid., p. 11). Empowerment evaluation is a response to a challenging idea for evaluation: that merit and worth are not static, and that within the

social system that programmes operate, nothing remains the same (ibid., p. 6). Populations and their knowledge shift and values and goals change correspondingly. Fetterman argues that internalised and institutionalised self-evaluation processes are necessary for evaluation to be responsive, and therefore useful, in the face of these changes (ibid.).

Fetterman considers empowerment evaluation adaptable to almost every environment (e.g. health, education, business), but as with Brunner and Guzman, he acknowledges the importance of commitment in order to undertake it. While management must clearly support the process and the risk-taking associated with it, groups themselves must request assistance rather than have this imposed upon them, and programme participants must want to control their own destiny and take charge of the specific steps required to do so (ibid., pp. 306–311).

Also, as with participatory evaluation, empowerment evaluation redefines the professional evaluator's role and their relationship with the target population to one of collaborator and facilitator rather than expert (Fetterman 1996, p. 5). Fetterman sees an evolving relationship between evaluator and programme participants and outlines five potential facets or developmental stages: training, facilitation, advocacy, illumination and liberation (ibid., p. 9). Through training evaluators teach people to conduct their own evaluations, demystifying the process and helping organisations to internalise evaluation practices. Because new skills are needed to respond to new levels of understanding, identifying what further training is required becomes an ongoing part of the self-assessment<sup>4</sup> (ibid., p. 11).

Facilitation evaluators are a step less directive than training evaluators, serving as coaches or facilitators to help others conduct a self-evaluation and providing guidance and direction to the effort. Fetterman also describes instances where evaluators serve as advocates for a group, notably in contexts where target groups are clearly disadvantaged. As he observes: *in an empowerment setting, advocate evaluators allow participants to shape the direction of the evaluation, suggest ideal solutions to their problems, and then take an active role in making social change happen* (ibid., p. 13). While all evaluations seek to reveal and provide information, Fetterman considers empowerment evaluation particularly illuminating of roles,

---

<sup>4</sup> Corresponding to Argyris and Schön's (1978) triple-loop learning, i.e. learning about learning.

structures and program dynamics. This sets the stage for liberation, i.e. freedom from these existing roles and patterns.

Fetterman (1994) describes four steps to help programme participants internalise evaluation as part of their programme planning and management. These are essentially: (i) take stock—which involves participants developing consensual scores of the strengths and weaknesses of components of the programme; (ii) establish goals – which also uses a consensual process to set priorities and agree measurable performance targets; (iii) develop strategies to accomplish goals; and (iv) document progress – which includes reaching an understanding of the type of evidence required to do this. Important overall themes in these steps are brainstorming, critical review and consensual agreement.

Both participatory evaluation and empowerment evaluation appear to blur the line between evaluator and social change agent. In fact it could be argued that social development workers should become more skilled in monitoring for programme improvement rather than expecting evaluators to take on roles of facilitation, collaboration building and empowerment. However, Fetterman, like Brunner and Guzman, and indeed numerous other advocates for more participatory approaches to evaluation (e.g. Papineau & Kiely 1996; Cousins & Whitmore 1998) assert the worth of participatory approaches within the evaluation community to counteract the way evaluation can misrepresent reality and actively disempower communities. To argue for these approaches they have needed to counter some common challenges. These include questions over the maintenance of research rigour, and range from straightforward doubts as to the objectivity of self-evaluation to concerns over the co-opting of the generation of knowledge by sub-power factions within the stakeholders you are seeking to empower. Those in support of more participatory approaches counter-argue that absolute objectivity is not feasible in evaluation and is overvalued in its contribution to the generation of useful knowledge in a typical evaluation context. They further assert that the collective contribution of multiple parties, playing a variety of roles within the programme, reduces the bias and counters overdominance by single groups (Fetterman 1996, p. 24).

Neither participatory evaluation nor empowerment evaluation is regarded as mutually exclusive with traditional evaluation methods. Thus neither offers the same fundamental critique to

traditional evaluation as that proposed by Guba and Lincoln (1989a, 1989b; Lincoln 1992) through what has been variously termed ‘responsive evaluation’, ‘naturalistic evaluation’ and ‘fourth generation evaluation’ (FG evaluation).

### 3.4.2 Construction of knowledge and meaning in evaluation

Guba and Lincoln (1989b) rest their proposal for FG evaluation on two platforms. Their first is similar concerns about equity and power to that which drove development of participatory and empowerment evaluation. As with these approaches, the focus of FG evaluation is the issues identified by the stakeholders, who they further define as being *persons or groups that are put at some risk by the evaluation* (ibid., p. 39). Since ‘knowledge is power’ and evaluation creates knowledge, Guba and Lincoln see stakeholders as open to exploitation, disempowerment and disenfranchisement through evaluation processes (ibid., p. 52). However, their second platform for FG evaluation is ontological rejection of previous positivist evaluation approaches in favour of constructivist methodology. This makes FG evaluation more reformist in its agenda than either of the previous evaluation approaches.

Constructivism has been discussed earlier in this thesis in relation to social learning (chapter 2) However, the replacement of scientific positivism with a framework based on constructivism is significant enough to make it useful to revisit some of the fundamentals and their implications for evaluation. In summary the distinctions are:

1. Constructivism implies relativist ontology, i.e. there are multiple socially constructed realities. This contrasts with positivist ontology based on the existence of a single objective reality.
2. Constructivism is characterised by a subjectivist epistemology that denies subject–object dualism. Thus it asserts that the inquirer and the inquired are interlocked. ‘Knowing something’ comes about through an interactive process by which the inquirer and the participant/respondent trade roles of teacher and learner.

The methodological response to these ontological and epistemological distinctions is to substitute the experimental approach to evaluation with what Guba and Lincoln describe as:

*...a hermeneutic dialectic process that takes full advantage, and account, of the observer/observed interaction to create a constructed reality that is as informed and sophisticated as it can be made at a particular point in time. (Guba & Lincoln 1989a, p. 44)*

Such a methodology includes context as part of the nature of the thing to be known (hence often referred to as *naturalistic*, i.e. taking place in natural, uncontrived settings) (Lincoln 1992, p. 7). It is 'hermeneutic' because it depends on continuing iterations of analysis and critique that lead to a negotiated and shared understanding of the situation (ibid.). A major task of the evaluator, therefore, is to conduct the evaluation in such a way that each group must confront and deal with the constructions of all the others. The theory being that...*their own constructions alter by virtue of becoming better informed and more sophisticated* (Guba & Lincoln 1989a, p. 41). Guba and Lincoln (1989a, p. 44), citing Guba (1987), list a number of assumptions underpinning constructivist-based evaluation. These include:

- Truth is not an objective reality but a matter of consensus among informed constructors.
- Facts have no meaning except within some value framework and hence there cannot be an objective assessment of any proposition.
- Phenomena can only be understood within context hence findings cannot be generalised from one situation to another.
- Evaluation produces data in which facts and values are inextricably linked. Valuing is therefore an essential part of the evaluation process providing the basis for attributed meaning.

From this come certain conclusions about evaluators. Fundamentally the evaluator in a constructivist-based inquiry is a partner with stakeholders in the creation of information. What distinguishes them from other stakeholders is their role as the organiser of the negotiation process that forms the basis of meaning making (ibid.).

Fourth-generation evaluation is described in terms of four general phases that may be reiterated or overlap (Guba & Lincoln 1989b, pp. 72–74)<sup>5</sup> (see Table 3.3). These are stakeholder engagement, collective review, information gathering, negotiation, and consensus. In both FG evaluation and empowerment evaluation, consensus building is an important theme. However, while FG evaluation tries to reach consensus on claims and issues, this is often not possible and conflict resolution is also a key component. In a departure from the more hands-off approach of participatory evaluation, the leadership role of establishing a conflict resolution process and moderating negotiation in FG evaluation falls to the evaluator.

Guba and Lincoln see the FG evaluation approach as able to address critical issues plaguing more traditional evaluation methodologies. Important among these are interrelated questions of how evaluations are used, the quality of evaluation information in decision making, and the association of evaluation and learning. In FG evaluation stakeholders are not only the direct users of evaluation information, but it is they, rather than the evaluators, who define what useful evaluation knowledge is. The use of a hermeneutic, dialectic methodology means that stakeholders are in a position to broaden the range of the evaluation inquiry (ibid., p. 53). Furthermore, engagement of a wide variety of stakeholders in the process focuses energy around those matters where there is disagreement and, in a process that gives balanced access to decision making, these can be the areas where most movement and improvement can be made (ibid., p. 54). FG evaluation is therefore ‘*responsive*’ both because it seeks the views of different stakeholders and because it responds to the most fundamental items in its subsequent processes (ibid., p. 41).

---

<sup>5</sup> Guba and Lincoln (1989, chapters 1 & 2) describe a four-phase process (p. 42), later further subdivided into nine steps (pp. 72–74). I have summarised these steps within the four phases first identified using my own summary headings.

**Table 3.3 Four phases in fourth generation evaluation** (Guba & Lincoln 1989b)

<b>Phase 1 Stakeholder engagement</b>	While the evaluation process remains open to new participants, it begins with a wide-reaching identification of the full array of stakeholders. During this stage claims, concerns and issues are identified.
<b>Phase 2 Collective review</b>	Identified claims are submitted to the collective for comment, refutation or agreement. The evaluation provides a methodology by which these various constructions can be understood, validated and ranked for importance. An agenda and negotiation process is prepared for any unresolved items.
<b>Phase 3 Information gathering:</b>	This phase involves collecting information to aid in addressing those issues and concerns that still emerge after phase 2. It may include training stakeholders to enable them to use this information.
<b>Phase 4 Negotiation and consensus</b>	A forum of stakeholder representatives is established and negotiation takes place under the guidance of the evaluator. The information collected is used and consensus is reached on disputed items. A mechanism for reporting findings is agreed. Those issues still unresolved go on to further cycles of evaluation.

Ultimately FG evaluation is about the mutual education of participants and therefore goes substantively beyond simply identifying views from stakeholders. Rather it asks them to confront different value positions and to become more aware of their own. While this is unlikely, even in the most ideal scenario, to result in an elimination of differences, it is also less likely to result in the kind of polarised debate over evaluation findings that frequently occurs in more traditional models, in which neither side expands their understanding (ibid., p. 56). Guba and Lincoln do not appear to regard FG evaluation as running the risks of rejection by professional evaluators based on lack of research rigour that participatory evaluation and empowerment evaluation encounter. In their view both traditional, positivist-based evaluation and this responsive constructivist evaluation are both forms of disciplined inquiry. This is because, in both, *the raw materials entering the argument and the logical processes they are subjected to* may be submitted for inspection (ibid., p. 44). However, they do anticipate resistance among the evaluation community in taking up the challenge of FG evaluation. Making the switch, they argue, to greater involvement of stakeholders is easy enough. Not so easy is the paradigm shift to a constructivist methodology for those researchers steeped in positivist science (and despite a steadily growing interest in constructivist responses to evaluation methodologies, hindsight would largely agree with them). Among the likely concerns for evaluators that Guba and Lincoln identify, one they share with other more

stakeholder centred evaluation approaches, is the reluctance to give up control to stakeholders, with all the methodological and political consequences this implies (ibid.).

However, in addition the constructivist basis to FG evaluation also implies uncertainty – both in the direction an evaluation might take and in the outcomes it might generate. This uncertainty, while uncomfortable for evaluators, is likely to be even less attractive to would-be commissioners and funders of evaluations. Equally unpalatable is the inescapable implication, (based on the inability to generalise from findings) that there are no universal solutions to problems. This clearly places the process of evaluation as more valuable than its outcomes in its capacity to generate learning and change, but limits its range to only that of the target group or programme it directly works with. However, this direct confrontation of ‘unknowability’ makes FG evaluation a good fit for the knowledge and learning demands inherent in social learning.

### **3.4.3 Putting theory into evaluation**

Not everyone sees Guba and Lincoln’s FG evaluation, or indeed any of the more stakeholder focussed approaches, as resolving all the critical issues in positivist, experimental evaluation. In particular a number of theorists have concerns about the absence of theory or the capacity for theory building in these frameworks, and it seems true to the behaviour of evaluation theorists that a number of different approaches in response to the ‘theory challenge’ have emerged. In this section I will look at theory-driven evaluation (Chen & Rossi 1989), and theory-based evaluation (Weiss 1991, 1995, 1997; Brickmayer & Weiss 2000; Connell & Klem 2000).

Stame (2004, p. 58) describes one of the fundamental challenges of social intervention programmes as *the black box problem*. What she is referring to is the reluctance of programme designers to openly examine what actually happens between the inputs of a programme (i.e. interventions) and the expected outcomes. In her view, too often programme designers pay little attention to how inputs or interventions are actually expected to work. Furthermore, she adds *evaluations do the same – concentrating on measuring outputs, whilst attributing the observed difference to the input* (ibid.). To explore the relationship between inputs and outputs i.e. to ‘open the black box’ requires an understanding of the theory basis for interventions. However, authors such as Chen and Rossi (1989, p. 301) note a pernicious failure of social intervention

programmes and their mainstream evaluations, to formally or explicitly specify theory. Stame (2004, p. 60) declares that not discussing programme theories amounts to *warranting programmes with 'absolute rationality'*, i.e. assuming that all needs are known and decision-makers are informed about the risks and opportunities implied in each option.

### **Theory-of-change approaches to evaluation**

In 1995 in a paper entitled 'Nothing as useful as good theory', Weiss presents the concept of grounding evaluation in theories of change<sup>6</sup>. Weiss's work rests on the idea that social programmes are underpinned by explicit or implicit theories about how and why the program will work. Furthermore, in practice, there are commonly several theories in action, which may or may not be consistent with one another. Even more sobering is Weiss's assertion that a common theory basis to a programme is simply '*K-A-P*', i.e. increased **k**nowledge leads to change in **a**ttitude which leads to change in **p**ractice (Weiss 1997, p. 510)<sup>7</sup>. The aim of theory-based evaluation is to examine the extent to which these theories hold; which assumptions break down, and where they break down; and which of several theories underpinning the programme are best supported by the evidence. In essence her theory-based evaluation works by surfacing theories, laying them out in detail, identifying all assumptions and then constructing methods for data collection and analysis to track their unfolding. Furthermore theory-based evaluation aims to describe the actual mechanisms that related to the desired outcomes, which may not necessarily be synonymous with the programme actions. For instance a programme action may be to 'pass on information' but the mechanisms for achieving a desired outcome may be the 'empowerment of gaining knowledge' (ibid.). Thus if a programme is successful in passing on information but the recipients are not empowered by this new knowledge, the programme can be said to have achieved its intended actions but still failed in its intended outcomes.

In Weiss's view theory-based evaluation serves four major purposes (1995, p. 4). Firstly, it concentrates evaluation attention and resources on the key aspects of the programme. This is

---

<sup>6</sup> Theory-of-change refers to the assumptions that underpin the steps leading to a long-term goal and the anticipated connection between programme activities and outcomes (Weiss 1995 in Anderson 2004).

<sup>7</sup> In my experience this simplistic formula of behaviour change is used as a basis for many environmental management programmes.

particularly useful in dealing with programmes and projects where there is complexity and uncertainty – even uncertainty about what is known and what is not, as characterises many environmental issues. Secondly, theory-based evaluation enables the aggregation of evaluation results into a broader base of theoretical and programme knowledge, i.e. it generates knowledge about key theories of change. This addresses some of the concerns raised about FG evaluation and its lack of applicability across programmes. It also points to the potential of evaluation contributing to the conceptual development of social learning.

Thirdly, theory-based evaluation requires that programme practitioners make their assumptions explicit and reach consensus with colleagues about what they are trying to do and why. This requires that programmes develop skills in dialogue and reflection that have the potential to foster the kind of ‘double loop learning’ that is one of the key components of social learning (see Figure 2.1 & Table 2.5). Articulation of theories and assumptions causes practitioners to confront their ideas about how they expect a programme to work and limits the possibility of different parts of the programme working at cross-purposes. Furthermore, in Weiss’ view, working through the *logic of their expectations* should inspire programme designers to consider more powerful interventions to achieve goals or to scale back expectations as they are forced to match them to the resources available (Weiss 1997, p. 517). Connell and Klem (2000, p. 116), writing about using a theory-of-change approach to education reform suggest that this approach helps participants to become more conscious and even more ambitious about the work they are doing, choosing *change-making work* over doing what is comfortable and familiar.

Connell and Klem (ibid.) identify four steps in a theory-of-change approach to evaluation: (i) generate or adopt an initial change framework; (ii) select indicators, populations, thresholds and timelines; (iii) develop action strategies to strengthen support for change; (iv) final implementation planning. Furthermore they outline a highly participatory process of engaging stakeholders in defining and exploring the theory of how and why an initiative works, and associate specific benefits for learning and development at each stage. For instance they suggest conversations about thresholds (step 2) forces stakeholders to consider the value of *what is enough*, which can change stakeholders mindsets about what needs to be done and what it is really going to take to do it. They argue that the high degree of dialogue and consensus required for this approach makes it more likely that programme stakeholders will follow through on commitments and persist together in overcoming setbacks.

The distinction between theory-driven (Chen & Rossi 1989) and theory-based (Weiss 1991) evaluation on the surface is not great and rests around some assumptions about theory and programmes. Chen and Rossi lament the atheoretical nature of the experimental paradigm dominating evaluation, observing that, without theory, programme goals are unclear and measures are consequently false (Stame 2004, p. 61). In contrast, Weiss comes to theory-based evaluation from a viewpoint that programmes are confused because of the complex reality of decision-making, which is a consequence of trying to act on multiple and competing theories. Bringing to light these theories-of-change enables a consensus to be developed on those that deserve to be tested. Thus the distinction between the two theory-oriented approaches is that while one places emphasis on locating suitable theories in an apparent absence of theory, the other concentrates on identifying, exposing, and reconciling the existing theories.

Connell and Klem (2000) take a highly pragmatic approach to the theory-oriented evaluation. In their view, the role of the theory is to outline the pathway of an initiative by making explicit both the outcomes (early, intermediate and longer term) and the action strategies that will lead to the achievement of these outcomes. The quality of the theory-of-change is judged by four criteria: (i) how plausible it is (i.e. stakeholders believe if you do this you will get the expected results); (ii) how doable it is in terms of human, political and economic resources; (iii) how testable it is; and (iv) how meaningful it is (i.e. stakeholders see the outcomes as important and worth the effort). However, they also recognise that the theory-of-change can itself alter as it is tested over the course of the initiative (*ibid.*, p. 94).

All these authors acknowledge a number of potential sources of theory. These include: (i) prior theory and research from the academic social sciences, (ii) exploratory research directed toward discovering underlying causal mechanisms of a programme, and (iii) extraction of the stakeholders' implicit program theory (Lipsey & Pollard 1989 in Chen & Rossi 1989, p. 301). All approaches also rest on discussion between evaluators and stakeholders to reach consensus on choosing a theory-driven approach that meets stakeholder needs (*ibid.*, p. 305). Connell and Klem (2000) observe that the value of an imported theory is that it enables discussions to progress away from self-protective explanations of what is currently done, and can attach the evaluation to a credible tested knowledge base. The downside is that people may not accept the imported theory and it may need a good process to enable people to engage with it.

Despite the considerable interest that theory-oriented approaches to evaluation have generated in the evaluation community, Weiss herself acknowledges a number of barriers to their use (1997, p. 502). Some of these, perhaps not surprisingly, centre on the notion of theory itself. Weiss (*ibid.*) suggests that the meaning of ‘program theory’ may be unclear and that a more useful terminology might be the word ‘*model*’ (although she acknowledges this is itself overused and subject to diverse interpretations). Further to this she notes a muddle amongst writers on theory-oriented approaches merging ‘theories of implementation’ (how a programme is being put into operation) with ‘theories of action’ (the anticipated chain of effects) (*ibid.*, p. 505)<sup>8</sup>.

There can also be difficulty with identifying or constructing programme theories, which may stem from a number of causes. In any programme there are multiple potential sources of theory, such as documents, prior research, programme funders, managers and participants. Interviews with the people involved to solicit information about theory may be challenging as the politics frequently surrounding programmes make it uncomfortable for them to produce ‘their theory’. Furthermore evaluators may find themselves faced with an articulated theory which they believe to be wrong and they are hence unwilling to rest their evaluation on it. A way around this might be to identify multiple theories and look at those that work and those that don’t. However, this adds to another surfacing problem with theory-based evaluation – the heavy demands on resources to undertake it (*ibid.*, p. 511). As if this was not all problematic enough, Weiss also warns that programmes don’t always go in the order of first establishing goals and then acting upon them. Goals frequently appear along the way. While this seems like it may call for an adaptive approach, Weiss sees this as adding an additional and too onerous a level of complexity (*ibid.*, p. 514). Finally, in her self-scarifying review of why theory-based evaluation may not be embraced as enthusiastically as she would hope, Weiss lists the demanding analysis, the inability to generalise from the results, and the likelihood that those evaluators who already do *process evaluation* may regard what they do as already close enough to theory-based evaluation.

---

<sup>8</sup> Together these two equate to a ‘theory-of change’ – see footnote 6.

Writing some time later Brickmayer & Weiss (2000) had more practice of theory-based evaluation to review. They still note that theory-based evaluation offers significant advantages for programme planning and improvement and the growth of knowledge around behaviour change, and offer this compromise:

*Even if the evaluators do not adopt the language of theory-based evaluation, they can incorporate elements of it into their studies...In the end, whether or not the theory is right it will have provided a framework for thinking about how the program is working...lead to development of creative ways to improve programs or design new approach and program theories – for future programs. (Brickmayer & Weiss 2000, p. 426)*

### **3.5 Tools and techniques for participatory developmental evaluation**

So far in this chapter I have concentrated on summarising developments in evaluation theory that are germane to using evaluation to build capacity for social learning. These have included approaches that are cognizant of the need to address power relations in evaluation; reorient the locus of the evaluation towards the interests of programme beneficiaries; focus on learning rather than information generation; and are both process and outcome oriented. It has also included approaches that have addressed the need for greater awareness and use of theory in programmes, as a means to explain events and to promote suitable actions in the future, and to build greater understanding about the overall system.

Regardless of which approach within this collective of participatory, developmental (P & D) evaluation theories a practising evaluator has most affinity for they are still left with the need to translate these ideas into an evaluation practice. In this they are faced with three particular challenges of method, of interest to understanding how evaluation can support social learning in environmental management: (i) how to help programme participants see across scale and systems; (ii) how to ask meaningful questions and facilitate a response to these; and (iii) how to promote reflection and dialogue.

### 3.5.1 Frameworks for seeing across scale and systems

While discrete programmes of activity with clear temporal and physical boundaries are still common, increasingly evaluations are sought to help multiple participants find their way through programmes of activities that span large and complex social and physical systems over lengthy periods of time, during which goals, objectives and methods of measuring them may change significantly. Even with the aforementioned discrete projects there is a growing interest in anchoring these within the context of the wider system within which they operate. Helping programme participants see across either the social, temporal or geographic system of their programme is all but impossible without some way of representing the system elements in a model or framework. There are broadly two purposes for frameworks (not mutually exclusive): (i) to reveal and improve the inherent sense of the programme, and (ii) to use theoretical and practice experience to illuminate some complex but essential ingredient of the programme.

An evaluation method emerging (or rather re-emerging since first appearing in the 1950s) that generates frameworks for improving programme management is the 'logic model'. Logic models are designed to create a picture of how a programme works by illuminating the underlying theory and assumptions, and highlighting how events are expected to unfold, what activities need to come before others, and how desired outcomes are achieved (W.K. Kellogg Foundation 2004). Advocates of the logic model approach cite three areas of benefit to programme development and evaluation. The first of these is improved programme design. This happens through keeping focus on outcomes, connecting interim to long-term goals, linking activities and processes to desired results, and keeping underlying programme assumptions at the forefront of the mind. The second benefit is providing the basis to ongoing programme evaluation. The logic model enables the programme participants to decide systematically which parts of programme activity to study, how to do so, and how to assess whether initial assumptions have been correct. Finally, the third benefit lies in the process of constructing the logic model itself, which brings stakeholders together to work on clarifying what is to be achieved, what they expect will be the outcomes, and what theoretical assumptions they are relying on. Changes are based on consensus and collective ideas rather than ideology or politics (ibid., pp. 35–40). Logic models thus incorporate participatory, constructivist and theory-oriented evaluation approaches.

Conducted with a strong eye to the process of participant engagement, logic models have the potential to surface tacit knowledge about the programme system and suggest causal relationships, and thus increase the collective knowledge about the programme's strengths and weaknesses. For instance a logic model can be used to interpret a bad outcome from a programme helping to clarify whether the results are a consequence of theory failure or some deficit in implementation. However, one of the complaints of the logic model approach is the tendency to invest large amounts of time deriving beautiful and explicit models of the programme that are then left without further reference to ongoing monitoring and learning. There is also a risk, inherent in the use of all models, that the model is perceived as 'the truth' rather than an explanation of the system based on the knowledge of all those who have contributed (Davidson 2008).

Frameworks can also be used in evaluation to introduce theories of how events happen that are not currently held by programme participants, thereby creating new lenses for participants to use to examine how their system is functioning. This is particularly relevant for those programmes which are grappling with complex systems about which much is unknown – a typical characteristic of programmes emerging in the arena of environmental management. Many frameworks are initially designed without evaluation in mind, rather as means to explain a complex system or set of relationships. However, any framework can form the starting point for an evaluation. The framework must be adapted and complemented with techniques that render it suitable as the basis for enquiry and learning. Therefore some process that tests and grounds the framework in the programme context must take place. Examples of using frameworks as a basis to evaluation are given in Chapter 6.

### **3.5.2 Ways of asking questions**

Asking questions is a primary tenet of evaluation. Not surprisingly then, the practice of specifying evaluation questions and developing questioning techniques has received considerable examination by the evaluation community. For evaluation that is designed to deliver programme accountability, questions that lead to specific measures of predefined criteria predominate. However authors such as Davidson (2008) regard questions that are prevalently descriptive in nature as *weak* – capturing what a client may want to know for immediate operational purposes but unable to come close to investigating important issues

about a programme. P & D evaluation, i.e. oriented towards generating knowledge at not only instrumental but conceptual levels, requires questioning that moves beyond data gathering or questions that can be satisfied by ‘yes’ or ‘no’ answers. Rather it requires questions that are open-ended, and depend on reflection, dialogue and the seeking out of additional information to fully comprehend (Preskill & Torres 1999). This is because questions not only elicit answers, but *good questions awaken the curiosity, challenge our limited views, and create opportunities for dynamic learning* (ibid., p. 97). In processes that are aimed at engaging groups or broad publics in actions for change, simply seeking out the galvanising question can be regarded as a most potent tool. As Chawla (1995 in Preskill & Torres 1999, p. 96) observes: *Asking the proper question is the central act of transformation.*

Garvin (1984 in Preskill & Torres 1999) offers a typology of questions to consider in establishing an evaluative enquiry, noting different types of questions can establish a fluid interchange, lead to surfacing assumptions, reveal disjuncts between theory and practice; or enable participants to think creatively about problems and solutions (Table 3.4). The questions in Garvin’s typology can all be used to promote discussion around what participants in a programme or members of an organisation already know and experience. However, as discussed above, models or theoretical frameworks that offer interpretations of systems can be a way of introducing new theories or ideas and questions based on these frameworks can therefore raise issues that would not otherwise be considered.

A checklist is one way of clustering questions together in a meaningful way. While a checklist in itself is quite a simple device it can be constructed on the basis of quite complex theory (Scriven 2007). It is an economical way of incorporating and presenting large amounts of information for stakeholders to engage with. Furthermore, by adding criteria of merit (e.g. scoring from 1 to 10) the checklist can be used as a basis to assessment (ibid.). Chapter 5 will look at the use of a checklist approach to evaluation.

**Table 3.4 Typology of questions** (from Garvin 1984 in Preskill & Torres 1999, p. 97)

Type of question	Examples
Broad diagnostic questions	What do you think the problem is?
Questions of action or decision	What would you normally do in this situation?
Questions of extension or synthesis	Exploring how different comments fit in with one another, e.g. how does that relate to what X said?
Questions of priority or ranking	What is the most important issue?
Questions of clarification	What do you mean by that?
Questions that challenge or test	Does this really hold true for every situation?
Factual questions	What? Commonly used but the least effective at sustaining dialogue
Hypothetical questions	If this situation was different how would it affect your decision?
Summary questions	What overall themes have emerged from this discussion?

Having determined the kinds of questioning approach that is most useful to employ, evaluators still commonly face several constraints. There can be ‘no-go’ areas of inquiry with restricted access, which can be explicit or camouflaged (the latter being where the evaluator finds they are simply blocked from ever having time with the people concerned). In contrast a constraint on questioning can also be the over-prescription by clients who specify areas of investigation based on a well meaning but under-informed mandate for the evaluation. In such a situation a role for the evaluator can be to widen the client’s view of the potential for the evaluation – to help *formulate a powerful and incisive set of ‘big picture’ questions to help guide the evaluation or clarify/modify an initial set developed by the client* (Davidson 2008, p. 2). Shifting the role of an evaluation relies on first developing a relationship with the evaluation commissioner that uncovers their more fundamental needs (an example of this is given in Chapter 5).

Even without client-imposed constraints on questioning there are few environments where active inquiry is innate. In organisations, solutions are frequently more highly valued than the questioning that led to them, and consequently are often embraced and acted upon without due consideration to their likely success. Questioning can be seen as time wasting at best but also

highly threatening. Exploring how to introduce an acceptance and even enthusiasm for questioning is examined in Chapter 7.

### 3.5.3 Reflection and dialogue in evaluation

Many authors have recognised the role of reflection in learning, and subsequently supporting and enabling reflection is emerging as a recognised methodological issue in evaluation. For instance Preskill and Torres (1999, p. 103) place reflection as one of the seven fundamentals of establishing an evaluative enquiry practice to support organisational learning stating:

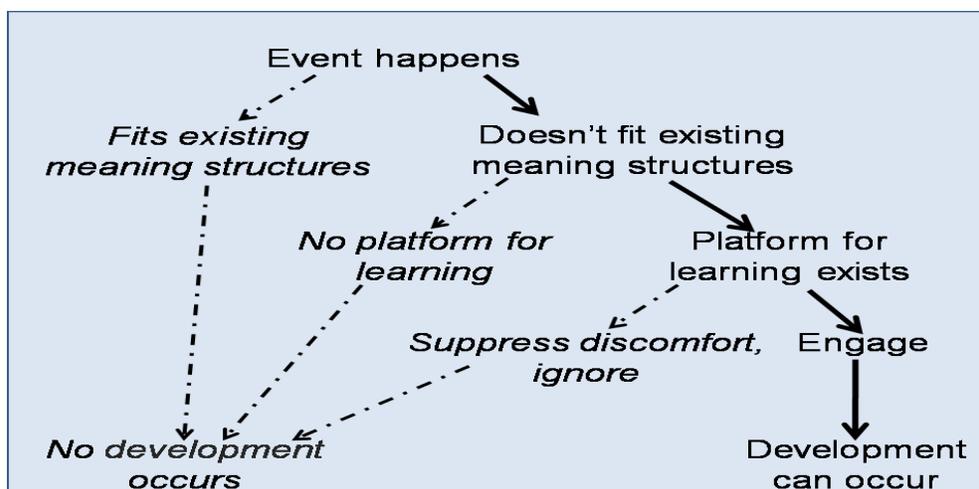
*[T]he reflection process is the way we come to know and understand ourselves. And, knowing ourselves is critical to creating new meanings that lead to personal development and change. Engaging in critical reflection as a group accomplishes an even stronger community of practice.*

To summarise an immense branch of literature *Not all experience educates* (Dewy 1938 in Merriam & Clark 1993, p. 41). The question then is what makes for good learning? While numerous authors have contributed substantively to understanding the basis to how experiences and new information are used to make and change meaning for individuals, most would recognise the central role that practices of reflection can have in this (Preskill & Torres 1999). Merriam and Heuter (1996, p. 251) summarise the role of reflection thus:

*For learning and eventually development to occur, we must engage ourselves with the experience. This engagement is a cognitive activity and may also have emotional and even physical dimensions. Most writers have labelled this step 'reflection', 'reflection – in-action' or 'critical reflection'.... We prefer the notion of 'engagement' which can include reflection before, during and/or after the experience; it can also include emotional and/or physical reactions.*

Mezirow (1991 in Preskill & Torres 1999, p. 101) suggests three foci of reflection: *Content* reflection – on the detail or description of a problem; *process* reflection – where critique is focused on the strategies for addressing the problem; and *premise* reflection – which considers the underlying assumptions or beliefs that underpin the problem situation. As identified in Chapter 2 (Figure 2.1, Table 2.5) the presence of platforms for reflection that address content, process and the underlying premise of the situation are fundamental to social learning.

Merriam and Heuter's (1996) theory of how an event can trigger development and the arrival at new meaning is summarised in Figure 3.2. It draws our attention to three criteria that are needed to create a transformative experience. The first is that the initial event (e.g. an action or a new piece of information) diverges from existing norms, values, and understanding. More particularly the event needs to be sufficiently different to the individual's existing mental model to present a challenge but not so different as to repulse them.



**Figure 3.2** Criteria for transformative experience (from Merriam & Heuter 1996).

Secondly, it is reliant on the existence of a platform or opportunity for learning. This needs to be safe, interesting and create some imperative to engage with the new experience or information so that the individual does not wish to ignore the opportunity but rather to engage with it (i.e. the third criteria). Fostering this environment is itself dependent on having available time, and supportive questioning that enables movement from a short-focus experience to its place in a wider context. For this to happen Merriam and Heuter (ibid.) suggest there is need for a 'journey guide' who assists the learner and who also models some aspect of the new territory. Chapters 5 and 7 will explore further the role of supporting reflection through evaluation.

Furthermore, reflective practice can have an impact not only on individuals but on groups as well. The basis to group reflection is dialogue, which, as discussed earlier in this chapter is a fundamental ingredient of FG evaluation. As Jarvis (1987 in Merriam & Heuter 1996, p. 247) observes, *it is possible for individuals to perceive what are apparently the same facts from a situation and experience them differently, even to experience them in such a manner as to*

*confer diametrically opposing meanings upon them.* Setting up an environment in which the exchange of views is possible introduces not only the potential for individuals to access new ideas but empowers the group with a diversity of opportunity.

As distinct from ‘discussion’ the purpose of dialogue is not necessarily to develop consensus, sell or convince others of ideas. Rather it is to share, enquire, and reveal. Dialogue contributes to learning where diversity of viewpoints is needed; and where the thinking around a problem situation has become stale, overgeneralised, dominated by particular views or values, or trapped in unrevealed assumptions about the situation. Like processes of reflection, the establishment of platforms for dialogue need careful attention. In particular, thought is needed on how to overcome problematic power dynamics or individual inhibitions that can stifle free exchange.

### **3.6 Summary: linking social learning and P & D evaluation**

In Chapter 2 I outlined how social learning emerged as a framework for understanding the social process demands inherent in the management of complex environmental issues.

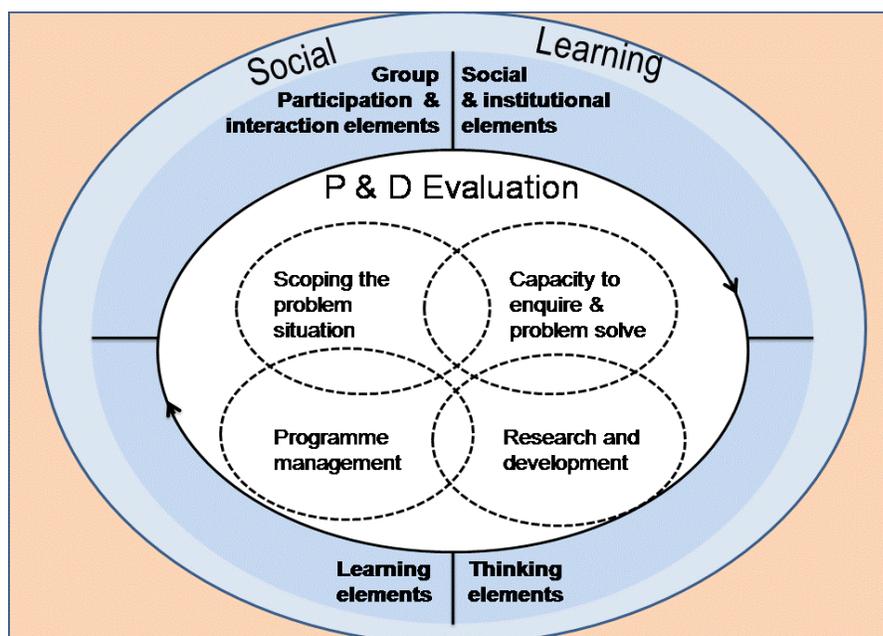
However, it is important to understand social learning not as a model for ‘how things should be done’ but rather as a set of premises or conditions – the management of which impacts on the ability of groups of stakeholders to find their way through complex problems where each share some knowledge, and towards which each need to take some action. The ideas that make up social learning are fundamentally about improving the conditions for learning and adaptation. There are no set steps to be followed, nor does it prescribe any particular starting position. Rather these ideas can be applied to improve the situation from ‘where you are now’. What social learning is reliant on then is the development of a culture and conditions for continuous and rigorous enquiry among the participants in the problem solving situation. This reflective practice examines not only what is known and needs to be known about the problem, but also what exists and needs to change about the social conditions in which the problem situation is located, i.e. learning about both content and process.

In search of a mechanism that might be used to drive this enquiry practice, in this chapter I explored a potential role for evaluation, as it might be applied to complex environmental problem solving situations. In particular I examine developments, in participatory-, reflection-

and theory-driven approaches which are able to be used to improve the learning capacity of groups, and to help environmental management programmes understand how they might go about working on social development aspects important to their overall goals. These evaluation approaches, tools and understanding around the creation of enquiry practice, offer a potential means to support the capacity for social learning in any given problem situation.

I propose four spheres where participatory, developmental evaluation approaches and social learning can intersect. These are:

1. Scoping the environmental management problem situation
2. Supporting the capacity to enquire and problem solve
3. Supporting the management of programmes or interventions in the problem situation
4. Research and development that facilitates the growth of theoretical and practical knowledge about addressing complex environmental management situations (see Figure 3.3.)



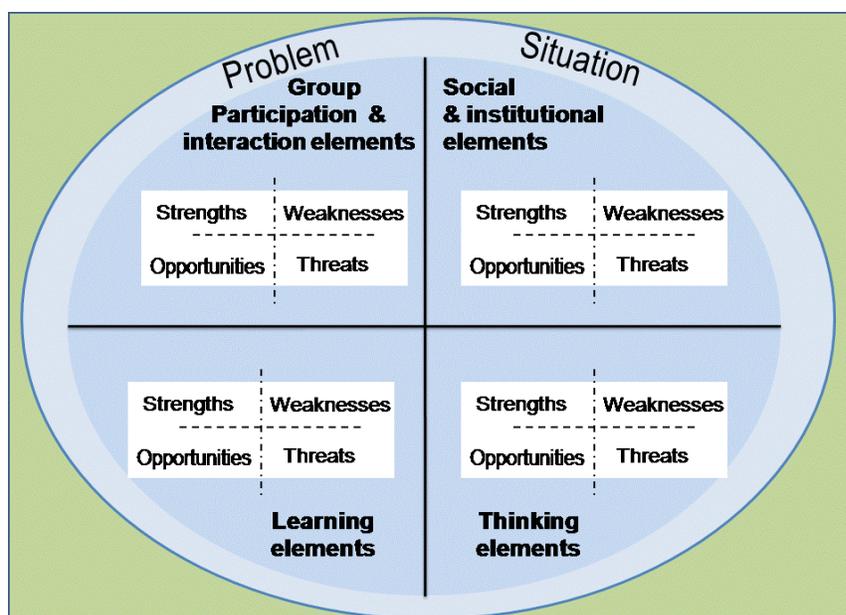
**Figure 3.3 Linking social learning and P & D evaluation.**

### **Scoping the problem situation**

Social learning, as a framework of ideas about complex environmental problem solving, can be used as the basis for a review of the problem situation as a whole. This includes first asking the question ‘Is improving the conditions for social learning important to addressing this problem?’

Only if the answer is ‘yes’ would you then look further into the major challenges, and opportunities and how they will present in this particular situation.

One way in which this could be done is by combining the Social Learning Framework with a basic SWOT analysis (strengths, weaknesses, opportunities and threats). A SWOT analysis of each of the four social learning quadrants as they apply to the particular situation can be used to analyse the overall social learning capacity of the situation (see Figure 3.4). For instance, a SWOT analysis of the capacity for participation and interaction – asking the question ‘How do existing platforms for conversations about this issue match up to the demands of multi-stakeholder collaboration and learning?’ might reveal existing networks that could be of value to the programme or potentially problematic relationships. Strategies to deal with these opportunities and threats can then be incorporated into the programme. Similarly a review of the social and institutional elements important to the situation may identify that existing decision-making arrangements are insufficiently flexible or open to the input of different stakeholders, which in turn might suggest other steps that need to be taken to change these arrangements or work around them.



**Figure 3.4** SWOT analyses of the social learning challenges of a problem situation.

In this way an evaluation based on the Social Learning Framework provides those tasked with driving improvements in the problem situation with a means to assess the existing social learning capacity and better consider both the challenges and opportunities for improvement. They may also use it to determine where their best rewards for effort might be found. Given the limitations on skills, resources and time available to most if not all complex environmental problem situations, it will be necessary for choices to be made on which areas are most amenable to change and which would yield the most strategic benefits. This is analogous to steps 1 and 2 of empowerment evaluation (Fetterman 1994), and an example of using frameworks to introduce social theory into programmes, and to support understanding across complex systems. The Social Learning Framework proposed in Chapter 2 was used as a basis for scoping the problem situation in each of the case stories presented in Chapters 4–7. A review of how this worked is included in discussion Chapter 8.

### **Supporting the capacity to enquire & problem solve**

P & D evaluation has many tools and approaches that are useful for improving the collective learning and problem solving ability of groups. Each of the cases presented in this thesis explores the use of one or more such approaches. In the Whaingaroa Catchment Management Project (Chapter 4), a participatory timeline generation approach was used to help the local catchment management group develop a sense of the successes and struggles of their work to date. In the Target Zero programme (Chapter 5), a checklist-based facilitated evaluation approach was used to improve the effectiveness of manufacturing company teams' as they worked to implement waste minimisation practices in their companies. In Chapter 6, I discuss the use of frameworks for describing the social processes of complex integrated environmental management programmes, and how a participatory evaluation based on one such framework (the social spaces framework) enabled the ICM programme to interpret competing demands for communication and information exchange. Finally, in Chapter 7 – Watershed Talk, a number approaches based on P & D evaluation were used to enable participants in the project to draw on their knowledge about their local environment, and discuss local intractable problem situations. These included techniques that encourage individual reflection, e.g. interviews, and photo-based story telling; and techniques that supported groups in their diagnosis of problems and consideration of options (e.g. soft-systems-based group reflection).

### Managing a programme

Supporting programme development and implementation is a traditional role for evaluation, P & D evaluation can be used proactively to plan programmes and interventions aimed at addressing an environmental management challenge (Table 3.5). A common evaluation approach, which combines P & D evaluation with theory-based evaluation to help design programmes that are ‘fit for purpose’, is the logic model (discussed in section 3.5.1). In the first instance, they can be used to help programme proponents, managers and participants clarify the logic of their programme, and uncover important assumptions and theories of action (or even an absence of theory) that underpin their approach. Using a consensual, participatory approach to developing programme logic can, at the least, assure all those concerned that they share an understanding about the programme’s intentions and direction. At other times what can be revealed are fundamental flaws, missing steps, or insufficient knowledge about how a part of the programme may work, and a decision can be made about how to address this.

**Table 3.5 Aspects of programme management that can be supported by evaluation**

Programme logic , assumptions, theories of action	Why are we doing this? What do we think will happen? How strong a link is there between actions and goals/intended outcomes?
Programme priorities	Given we expect things to happen in this way, what is the most useful thing to do? What can we afford to take a risk on?
Programme implementation	What can we use to gauge what is happening (indicators)? How will we gather information on these indicators and assess what they are telling us (monitoring strategy)?

By exposing the link between proposed actions and intended outcomes, P & D evaluation can also be used to help programmes allocate priorities for resources and effort, by clarifying what actions are most closely linked to the most important intended outcomes. It can also be used determine indicators that are appropriate and practical to use as a means to gauge programme progress; as well as establish a plan for gathering information on these indicators, and schedule for reviewing the meaning of these indicators for the programme’s ongoing operation. Unanticipated developments can occur in the best-planned programmes, and monitoring ‘for surprise’ (Guijt 2008) enables programme proponents, managers and participants to make choices about how they respond to these ‘windfalls’ or ‘pitfalls’.

Using a logic model approach to provide overall guidance to the development of a particular programme may at times highlight the need for a more in depth information need in particular areas which other P & D evaluation methods might assist with. For example stakeholder analysis can be used to identify a project's key stakeholders, assess their interests and needs, and clarify how these may affect the project's viability (Allen & Kilvington 2009). Stakeholder analysis might also be combined with the Social Learning Framework to reveal needs and abilities of stakeholders that will contribute to the social learning capacity of the situation. For instance, the stakeholder analysis might be designed to explore the role of particular stakeholders in influencing institutional arrangements or isolate their particular potential knowledge contribution to understanding the problem system. Similarly, where programmes of work extend across large geographic and temporal scales, various framework-based evaluations may be applied to help programme participants and managers visualise the more intangible social and/or physical changes that are underway.

### **Research and development**

The final area in which P & D evaluation and social learning intersect to support complex environmental problem solving is through research and development. This is far from an academic afterthought. Rather it enables environmental management programmes to become active contributors to the body of knowledge on complex environmental problem solving that supports their own practice. Indeed the cases in this thesis are all examples where practitioners, as much as researchers, have an active interest not only in learning whether their programme is working but in understanding the reasons why, to enable these practices to be more thoughtfully and successfully applied elsewhere. An example of where evaluation played an critical role in research and development of new environmental management practice is the Twin Streams project, conceived of and sponsored by the Waitakere City Council. Ostensibly aimed at improving water quality within an urban catchment in the Auckland Region, this project took a departure from the conventional, agency-centric, approach by using a community-development-driven methodology. Testing the assumptions underlying this approach (i.e. was this really going to make a difference?), and clarifying the achievements and limitations of the new way of working were regarded as essential roles for the programme evaluation (Chilcott, personal communication, October 2009. See Appendix 1). Important to choosing the right evaluation

approach to support research and development, is considering ‘who is the learning for?’ It may be that in a complex initiative there are multiple learning needs operating at different levels.

### **3.6.1 Putting evaluation into practice to support social learning**

This chapter reviewed several strands of evaluation approaches, tools and methods useful to building capacity for social learning. Importantly though none of the explored approaches is sufficient in itself. For instance while it is easy to imagine that participatory evaluation can be successful on a project basis, i.e. improving the ability of project participants to have influence on project directions, how does it perform in broader decision-making contexts when a mixture of programme participants and non-participants are responsible for the direction of the initiative? Theory-based evaluation takes away the problems of ‘self-assessment’ and ‘culture capture’ perceived to be problematic in participatory approaches, but without participatory and empowerment-based methodologies, it lacks the potential for developing ongoing capacity for self-assessment and learning, or the constructivist emphasis that facilitates the dialogic processes essential for communicative rationality. Fortunately, although there is undoubtedly a degree of ‘adherence to faith’, modern evaluation practice does not require us to make choices between these different evaluation theories.

In addition to the part specific approaches to evaluation can play in building capacity for social learning, there is a role evaluation can play in itself – that of situating inquiry within a valid social and institutional setting. In an age in which monitoring, measurement and accountability is all pervasive, an observed impact of the ‘pernicious audit’ is that, that which can be most easily assessed becomes the focus of endeavour (Shore & Wright 1999). As evaluation practice itself transitions from its historical accounting orientation to focus on learning and stakeholder empowerment, the potential is to counter cultural bias and introduce capacity for social learning by the back-door.

What cannot be avoided is the influence of previous experiences of evaluation. These are likely to be jaded, as for many, evaluation is seen as taking time and resources, distracts from core work, and is commonly associated with a kind of ‘pass or fail’ decision-making by an external agent. In addition, in the messy world of actually doing evaluation, it is questionable whether

people really want you to warn them of trouble ahead (Cronbach et al. 1980). Clearly the further back an evaluation can be located (i.e. into programme development rather than simply programme outcome), the more successful an evaluation intervention can be in developing capacity for improvement and learning, and the more optimistically it is likely to be regarded. Ultimately, it is important to note that, although the theory and practice of evaluation can offer much to the challenge of building capacity for social learning, the choice of approach will often depend on available resources and the recognised mandate for the work.

The cases presented in Chapters 4–7 are about three environmental management programmes set up to address a particular set of issues, and tell the story of the application of P & D evaluation approaches designed to improve some aspects of the social learning potential of a given situation. In each instance these interventions are imperfect, and opportunistic, but deal with the anticipated and unexpected outcomes of the experience, the barriers and the pitfalls.

Each of these case stories follows a schema of questions (see Chapter 1, Table 1.1, repeated in Box 3.1). The cases begin with a review of the critical factors that frame the social learning challenge of the situation. This includes the particular social learning capacity needs and the match between these and the programme intervention that is occurring. This is essentially a SWOT analysis based on the Social Learning Framework described in section 3.6.

**Box 3.1 Schema of questions for case studies**

- *What is the social learning challenge of the situation?*
- *What aspect of social learning was supported by the evaluation?*
- *What evaluation approach was chosen?*
- *What happened/results/outcomes?*
- *What was learnt?*
- *What is the significance of this?*

The case story then outlines how evaluation was designed to contribute to building capacity for social learning; what evaluation approach was used and how it was implemented; and the outcomes from this

(intended and unintended). Finally it considers what was learnt from the use of evaluation in this situation and the significance of this for understanding how evaluation can support social learning in environmental management.