TOWARDS A NEW FRAMEWORK FOR EVALUATING SYSTEMIC AND PARTICIPATIVE METHODS

Gerald Midgley, Jeff Foote, Annabel Ahuriri-Driscoll and David Wood
Institute of Environmental Science and Research (ESR) Ltd., 27 Creyke Road, PO Box 29-181, Christchurch, New Zealand.

ABSTRACT

Systems practitioners often make significant claims for the value of their methodologies and methods. However, when evidence is presented to support these claims, it is usually based solely on the practitioner's own reflections on single case studies. Less often, practitioners set up post-intervention debriefings with project participants using questionnaires. While the latter is an improvement on researcher reflections alone, there have been few attempts at systematically evaluating across methods and across case studies undertaken by different practitioners. This is understandable because, in any given local intervention, contextual factors, the skills of the practitioner and the purposes being pursued by stakeholders are inevitably going to affect the perceived success or failure of a method. The use of standard metrics and even qualitative criteria for comparison can therefore be made problematic by the need to consider what is unique in each intervention. So is it possible to develop a single evaluation approach that can support both locally meaningful evaluations *and* longer-term comparisons between methods? This paper offers a framework for the evaluation of methods that seeks to do just this. Research on the framework and associated tools is in its infancy, but pilot studies suggest that it is promising. Comparing across methods will ultimately require the development of a longer-term international research program, and the paper serves as a first call for participants in this.

Keywords: evaluation of methods, methodology, participative methods, systems methods.

EVIDENCE FOR THE VALUE OF SYSTEMIC AND PARTICIPATIVE METHODS

The literature contains many methodologies and methods to support systemic and participative change. When claims are made for their success or failure in practice, the authors making those claims are usually required to justify them. Various reviews of the literature on the evaluation of systemic and participative methods suggest that most of the justifications provided by practitioners are based on personal reflections alone (Entwistle et al, 1999; Connell, 2001; Rowe and Frewer, 2004; Sieber, 2006; White, 2006). Clearly many practitioners are highly experienced, so their reflections should not be dismissed out of hand. Nevertheless, unless they think broadly and from different perspectives about the criteria they use to evaluate their interventions, they may miss evidence that does not fit their current thinking about what is important (Romm, 1996). We therefore suggest that there is a need for caution in accepting practitioner reflections alone as reliable evidence of success or failure.

Most practitioners undertaking evaluations beyond personal reflections tend to conduct post-intervention debriefings or interviews with project participants. These evaluations are often based on explicit criteria reflecting the practitioner's experience, a given theory, a literature review and/or stakeholder expectations generated through a consultative exercise (Beierle and Konisky, 2000; Rowe and Frewer, 2004). In some cases, formal evaluation instruments have been developed and applied (e.g, Duram and Brown, 1998; Rowe et al, 2004; Berry et al, 2006). Also a number of practitioners advocate triangulating across two or more evaluation methods such as structured, semi-structured and unstructured interviews; focus groups; participant observations; surveys; literature reviews; and document analyses (Duram and Brown, 1998; Buysse et al, 1999; Charnley and Engelbert, 2005; Rowe et al, 2005; Cole, 2006; McGurk et al, 2006).

What is clear from the literature, however, is that only a very small minority of studies seek to compare between methods or across case studies undertaken by different practitioners (Halvorsen, 2001; Rowe and Frewer, 2004). A notable exception is Beierle and Cayford (2002) who quantitatively compared broad classes of method using a standard set of variables applied to 239 case studies of public participation and concluded that more intensive processes (such as mediation workshops) are better than less intensive processes (such as public meetings) at achieving a wide range of outcomes. We suggest that systemic problem structuring workshops are relatively intensive compared with several of the processes investigated by Beierle and Cayford (2002), so this gives us grounds to be cautiously optimistic about the value of systemic and participative practice. However, we cannot take this study as strong evidence because they did not specifically identify systemic problem structuring workshops as a category for comparison with other approaches.

The overall picture is therefore of many claims for the benefits of a diverse array of systemic and participative methods, with varying degrees of evidence provided by practitioners to support these. Only a few studies have compared across methods, and even these have only been able to contrast broad classes of approach.

The key question is: what kind of evaluation is both necessary and possible? We have already argued that practitioner reflections alone can be problematic, but are there methodological or practical reasons to prefer either locally focused evaluations (possibly with some learning across case studies, when this is feasible) or large-scale, quantitative comparisons between methods?

Different Evaluation Approaches

Rowe and Frewer (2004) classify approaches to the evaluation of methods into three types. First there are '*universal'* evaluations: i.e., ones claiming to produce knowledge that is applicable across all types of method and intervention. According to Rowe and Frewer, to achieve universality, large-scale quantitative studies are needed. Nevertheless, to make comparisons possible, only variables of general relevance across all methods and interventions can reasonably be assessed. Next there are *local* evaluations: comparing between a subgroup of methods or intervention types. These

require smaller scale studies and can incorporate more detailed questioning, as the variables to be examined may be relevant only to the subgroup of methods under study rather than to all possible methods. Some researchers working on local evaluations advocate a quasi-experimental approach, either testing methods in the laboratory or in controlled field conditions. Rowe and Frewer (2004) call the third and final type of evaluation, which the majority of practitioners use, *specific*. This means focusing on only one method or intervention. The advantage of this is that the evaluation can be made locally relevant, drawing (for example) on information about the unique expectations of stakeholders to establish evaluation criteria. Rowe and Frewer argue that, while it is difficult (for practical reasons) to conduct truly universal evaluations, practitioners and researchers should aim to achieve as much generality as possible, and should certainly do more than undertake evaluations with only a specific remit because generalisations that are made from the latter are scientifically unfounded.

However, for both epistemological and methodological reasons, we do not accept that it is possible to generate universally applicable knowledge about methods. Our epistemological argument is that knowledge (or understanding) is always linked to the purposes and values of those producing or using it, and is dependent on the boundary judgements they make (Churchman, 1970; Ulrich, 1983; Alrøe, 2000; Midgley, 2000). To claim that knowledge about systemic and participative methods (or any other phenomenon for that matter) is universal is to ignore the purposes, values and boundary judgements that make the knowledge relevant and adequate for a particular context.

We also have two methodological arguments following from our epistemological one. First, claiming universality for knowledge about systemic and participative methods would suggest that this knowledge will remain stable over time. However, it is clear from the literature (e.g., Shaw et al, 2006; Franco et al, 2007) that new systemic and participative methods are being produced on a regular basis, indicating that people are learning from previous practice and are also having to respond to an ever increasing number of unique practical situations. Given that this is a dynamic research environment, it would seem risky to assume that a standard set of variables will always be relevant. Undertaking a series of more limited comparisons between particular methods might be methodologically wiser than trying to set up a 'universal' study.

Our second methodological argument is that only seeking knowledge about the supposedly generic strengths and weaknesses of methods ignores legitimate questions that can be asked about the effectiveness of those methods in particular local circumstances. Given that systems practitioners work most of the time in particular contexts with unique features, it would only meet a small fraction of the need for evaluation if we were to ignore non-generic questions, and this would be unacceptable to local stakeholders wanting to know what will best meet their particular needs.

There can also be problems on occasion with what Rowe and Frewer (2004) call 'local' evaluations: comparing more limited sets of methods in smaller scale research projects. Some have called for 'objective' local studies rather than the simple reporting

of subjective impressions (Rowe and Frewer, 2004; Rowe et al, 2005). However, when the pursuit of objectivity involves a retreat into the laboratory to conduct controlled experiments, then the validity of the comparison of methods has to be questioned due to the artificiality of the situation (Eden, 1995; Shaw, 2003; White, 2006). If quasi-experiments are established in the field, then this raises other problems: McAllister (1999) argues that it is unethical to use a control when dealing with real community issues, and Duignan and Casswell (1989) simply point to the impracticality of finding two situations that are sufficiently alike to make a comparative study robust.

In making criticisms of attempts to take a controlled or quasi-experimental approach, some authors have advanced alternatives. Kelly and Van Vlaenderen (1995), Jenkins and Bennett (1999) and Allsop and Taket (2003) adopt an action research or 'emergent' methodology: i.e., one where criteria for evaluation emerge through engagement with stakeholders. Eden (1995) makes the important point that most interventions are complex, and practitioners can rarely anticipate everything that will become important, so the evaluation approach needs to be able to respond to the unexpected.

However, does this mean that evaluations cannot legitimately generalise from single, specific case studies to other contexts that may be similar in at least some respects? It is certainly true that the 'success' or 'failure' of a method in any particular case results from use of the method-in-context and cannot be attributed to the method alone (Checkland and Scholes, 1990; Buysse et al, 1999; McAllister, 1999; Murphy-Berman et al, 2000; Morgan, 2001; Margerum, 2002; Rowe and Frewer, 2000, 2004; Branch and Bradbury, 2006; McGurk et al, 2006; White, 2006; Warburton et al, 2007). Nevertheless, several action researchers claim that *cross case study learning* is possible, with practitioners reflecting on similarities and differences between cases (e.g., McAllister, 1999; Yearley, 2006; White, 2006). Checkland (1981) argues that evaluating a systems methodology depends on the long term accumulation of evidence from a diverse range of applications, giving progressively more confidence that the approach is useful across contexts.

A Pragmatic Step Sideways

It would appear from the literature that most practitioners accept the above logic and are more inclined to undertake specific, locally meaningful evaluations (and possibly learn across these) than attempt comparisons between methods using generic, quantitative measures (Mingers and Rosenhead, 2004; White, 2006). However, we have to ask whether this means that all forms of quantitative comparison are redundant. White (2006) argues that the debate has become unhelpfully polarised, with many advocates on both sides taking 'purist' positions and spurning methods that could enhance their own evaluation practices. He therefore proposes a more pragmatic line: identifying important research questions and asking what evaluation methods might answer these most effectively. We agree that this is a useful step sideways from the either/or debate, but we nevertheless suggest that identifying effective evaluation methods to address particular research questions involves

considering the *practicalities* of undertaking evaluations as well as the norms of what constitutes a valid or legitimate methodology. A difficult balance has to be struck between rigour and relevance (Shaw, 1999) because if the former is unquestioningly prioritised then there is good evidence that stakeholders will not co-operate (Rowe et al, 2005). Importantly, this balance has to be struck regardless of whether an emergent action research approach is being followed or whether a more traditional scientific study comparing methods is being undertaken.

In sympathy with White's (2006) pragmatic intent, we set out to propose an evaluation approach that supports locally meaningful evaluations *and* is capable of generating data for longer-term quantitative comparisons between methods without compromising local relevance. The overall framework is based in the tradition of systemic action research, but instruments can be employed as part of the emergent evaluation of methods that enable data gathering for both immediate local and longer-term comparative use. Below, we outline the rationale for our framework. We then discuss early work in developing and testing a questionnaire that can be used in the context of it. We end with a call for international participation in a long-term research program to improve both one-off local evaluations *and* compare between methods using a set of generic criteria.

A NEW EVALUATION FRAMEWORK

Our evaluation framework is represented in Figure 1. An evaluation using it is primarily focused on the use of a particular *method* (or set of methods) in a *context* for particular *purposes*. The words in italics in the previous sentence represent what we regard as three necessary aspects to evaluative inquiry, and they need to be interrelated in the context of a specific reflection on the use of a method. Exploration of these aspects may proceed in any direction around the circle in Figure 1, and may loop back and forth according to the needs of those involved in the evaluation.

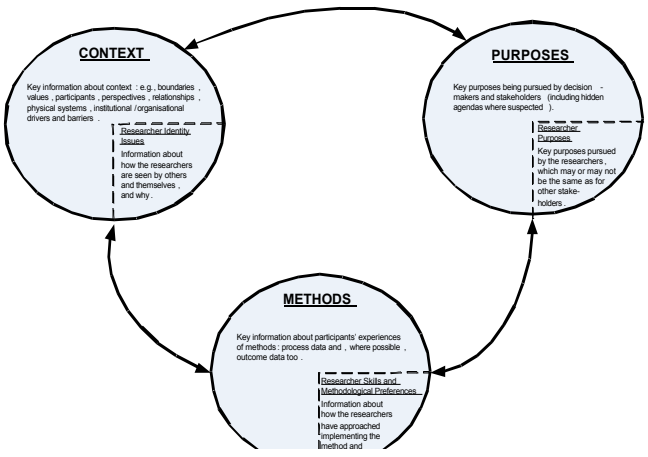

Figure 1. Framework for the Evaluation of Methods

Other authors have proposed similar, but not identical, frameworks. Buysse et al (1999) and McAllister (1999) advocate the exploration of both purposes and context,

but tend to take as given the nature of the method to be evaluated. Flood (1995) asks practitioners to reflect on the adequacy of their contextual analyses, their choices of methods and their intervention outcomes. Similarly, McGurk et al (2006) look at methods in light of the context and their outcomes. Here, however, the purposes being pursued become implicit: whether or not these differ from the outcomes is not necessarily at issue. As far as we are aware, the only authors to have proposed the same three interrelated aspects as us (context, purposes and methods) are Warburton et al (2007), although they do not consider the implications of the practitioner's role in the situation. This is an important issue for us (and is represented in Figure 1 by the text in the bottom right corners of the three circles) because our experience is that the practitioner becomes an *interactive part* of the situation in which s/he is seeking to intervene using systemic and/or participative methods (Checkland, 1981), and his/her identity and relationships can significantly effect the trajectory of an intervention (Mingers, 1997; Midgley et al, 2007).

In our approach, when looking at a single case study, there is no pretence that it is possible to evaluate a method independently from the purposes it is put to and the context in which it is applied. Nevertheless, we can still inquire about the *relationships between* the method, purposes and context. Inquiry focused on an intervention can look at how satisfactorily the method addressed given purposes, and what aspects of the context enabled or constrained its application. Some features of the context-purposes-method relationship may be apparent early on in an intervention, while others may only emerge as the inquiry unfolds. Hence the utility of an action research approach for the evaluation of methods, which remains open to the emergence of new understandings as inquiry deepens (e.g., Kelly and Van Vlaenderen, 1995; Jenkins and Bennett, 1999; Allsop and Taket, 2003).

Below, we examine the three aspects of evaluation (context, purposes and methods) in turn, explaining why each of these is important to developing a rounded understanding of how a method has operated in a particular case study of practice.

Context

More has been written about context than the other two aspects of evaluation, arguably because it is crucial to good practice to realise that the same method utilised by the same practitioner can succeed or fail depending on the complexities and dynamics of the situation (e.g., Checkland and Scholes, 1990; Buysse et al, 1999; McAllister, 1999; Murphy-Berman et al, 2000; Rowe and Frewer, 2000, 2004; Morgan, 2001; McGurk et al, 2006; White, 2006; Warburton et al, 2007).

Relevant aspects of context identified by Jackson and Keys (1984) are the complexity of the issue being addressed using a systems approach and the relationships between the participants. In contrast, Margerum (2002) identifies potential contextual inhibitors of effective participation: a low level of commitment by key decision makers; parochialism (which can negatively affect inclusiveness); participants having inadequate skills and abilities; operational issues preventing the implementation of ideas; a lack of strategic thinking beyond the exercise at hand; poor leadership; and

scarcity of resources. Ong (2000) discusses the facilitative effects of strong social capital, and Alberts (2007) documents the negative effects of participant inexperience and ignorance of technical issues. Branch and Bradbury (2006) claim that a key aspect of context is 'managerial' attitude: especially the disclosure (or not) of relevant information; whether managers set agendas unilaterally or are open to power sharing; whether or not there is mutual respect in relationships; whether there is accountability to stakeholders; and whether or not people believe that a transparent decision making process will be used following stakeholder participation. Kelly and Van Vlaenderen (1995) concentrate on stakeholder interactions, looking at how patterns of mistrust and miscommunication can become established and affect the use of participative methods. Related to this is the identity of the practitioner: Midgley et al (2007) discuss how identity issues can make a significant difference to the quality of relationships, and hence the success or failure of a method (this is represented in Figure 1 by the bottom right hand corner of the 'context' circle).

No doubt the list of possible aspects of context could be extended indefinitely, and different issues will be relevant in different situations, so it is arguably more useful to give some methodological guidelines for exploring context in local situations than it is to provide a generic inventory of variables. We suggest that the following guidelines, derived from reflections on different systems paradigms (as represented by Jackson, 1991, and others), can all contribute in different ways to *boundary critique* (the exploration of different possible boundaries, or frames, for a contextual analysis):

- Underpinning different boundary judgements may be quite different perspectives on the nature of the context (Churchman, 1970). Therefore, exploring diverse perspectives (e.g., as advocated by Checkland, 1981) may lead to the identification of alternative possible ways of bounding a contextual analysis (Ulrich, 1983).
- Establishing a boundary for analysis involves making a value judgement on what issues and stakeholders are important or peripheral (Ulrich, 1983). Therefore, undertaking an exploration of different stakeholders' values and priorities can be helpful. It is also useful to identify conflicts between people making different value judgements as well as processes of marginalisation that may constrain stakeholder participation or make the discussion of some phenomena taboo (e.g., Midgley, 2000).
- Identifying the presence of influential institutional or organisational systems may be important. Any such system can have its own agenda, rationality and momentum that may come to dominate an intervention (Douglas, 1986; Luhmann, 1986), yet organisational systems still have to interact with others, and tensions can result. Thus, an institutional analysis can be a useful aspect of boundary critique.
- There may be socio-economic and ecological systems providing resources that can be used constructively by participants, or these systems may impose limits on what is achievable without incurring negative side-effects (Clayton and Radcliffe, 1996). Economic issues may point to concerns about social justice, which (if present) could influence people's perceptions of the effects of systemic and

participative methods: i.e., the use of a particular method may be seen as supportive of just or unjust social relationships (Jackson, 1991), so it can be useful to look at the effects of socio-economic systems as part of boundary critique. Taking explicit account of ecological systems can also enhance boundary critique by challenging a Western cultural tendency to uncritically resort to boundaries defining exclusively human systems, thereby marginalising the ecological (Midgley, 1994).

- Within and across ecological, economic, social and organisational systems, there may be important causal pathways, and in particular feedback loops, that can point to systemic enablers of, or constraints on, an intervention (e.g., Forrester, 1969). Bateson (1970) argues that it is important not to 'cut' relevant feedback loops, and again this is a good principle to inform boundary critique: when we see interconnections stretching beyond people's usual understandings of context we can ask whether it is important to widen the boundaries of analysis to account for these.

Essentially then, a useful approach to exploring context may involve looking at different possible boundaries for analysis, concentrating in particular on different stakeholder perspectives; value judgements around the inclusion or exclusion of issues and stakeholders; processes of conflict and marginalisation; ecological, economic, social and institutional/organisational systems that may act as enablers or constraints; and causal relationships and feedback processes within and across those systems.

Purposes

The second aspect of our evaluation framework is concerned with exploring stakeholders' *purposes* in engaging with an intervention. Purposes are closely linked with values and motivations (McAllister, 1999), and they are important to an evaluation because particular methods are likely to appear more or less useful depending on the purposes being pursued. Different methods are generally good for different things (Flood and Jackson, 1991), and it is the perceived 'fit' between purpose and method that is important to evaluate: a disjunction may be responsible for an attribution of failure.

It is important to consider possible hidden agendas as well as explicitly articulated purposes. These may significantly affect the trajectory of an intervention (for instance through sabotage), and thereby the evaluation of the method used (Ho, 1997). It is also useful to look out for mismatches between articulated purposes and ones attributed by others (both to individuals and organisations) because mismatches of this kind often signal mistrust or conflict that will be relevant to the performance and evaluation of methods (Kelly and Van Vlaenderen, 1995).

Whether or not there is mistrust or conflict, there will often be multiple purposes at play. If people come to an intervention with different purposes for engaging, then it is likely that different evaluation criteria will be important to them (McAllister, 1999; Murphy-Berman et al, 2000; Tuler et al, 2005; Rowe and Frewer, 2004; Masozera et al, 2006; White, 2006). While Rowe and Frewer (2004) say that an appropriate

response is to *set aside* the purposes and preferred criteria of diverse stakeholders in favour of a single criterion of 'acceptability of the method to all parties', more nuanced findings will be generated by evaluating the method against multiple criteria of relevance to different stakeholders (Murphy-Berman et al, 2000).

Note here that the purposes of the practitioner should not be excluded from consideration. There may be a good 'fit' between stakeholder and practitioner purposes, but there may also be disjunctions. An example is when the practitioner is a university researcher and brings a pre-defined academic research agenda into the intervention, which may influence how systemic and participative methods are chosen and used. Even when an academic practitioner makes a significant effort to be responsive to stakeholders, there may still be mistrust stemming from *expectations* of divergent purposes (Adams and McCullough, 2003), and this may affect the evaluation of methods.

Methods

Earlier we mentioned that some authors have advocated taking account of the effects of stakeholder purposes and context, but they tend to take the nature of the method being evaluated for granted. It is important not to do this because different methods make different theoretical and methodological assumptions about (amongst other things) human relationships, knowledge, and the nature of the situation that is the focus of the intervention (e.g., Jackson, 1991; Romm, 1996; Midgley, 2000). In an evaluation, we need to be able to account for if and how these assumptions have shaped the unfolding of the intervention.

There may also be elements of methods that people in some cultures (or sub-cultures) will find it easier to accept or work with than others. While culture may be conceived as an aspect of the context, it may also be reflected in the construction of a method, which is why a number of methodologists working outside the Western tradition have sought to establish systems and other approaches developed from their own philosophical and cultural perspectives (e.g., Smith, 1999; Zhu, 2000; Shen and Midgley, 2007). Becoming aware of the cultural norms embodied in a method may be important to understanding its effects across cultural contexts.

The *process of application* of a method is important as well, not just the method as formally constructed. For instance, the same basic method may be enacted in quite different ways depending on the preferences and skills of the practitioner and the demands of the situation at hand. Compare, for example, two significantly different accounts of soft systems methodology (SSM): Checkland and Scholes (1990) discuss how the methods from SSM should be utilised in a flexible and iterative manner, while Li and Zheng (1995) insert some of the same methods into a 'general systems methodology'. In the latter case, it is clear that the methods of SSM are to be applied in a linear sequence. In many contexts, such a significant difference in the process of application of the same set of methods is bound to impact upon the way the method is perceived.

Not only can the practitioner's preferences and approach be important, but also the extent of his/her skills and experience may influence whether the use of a method is perceived as successful or not. Mingers (1997) describes these as the "intellectual resources" that the practitioner brings into an intervention, and it is important to be able to distinguish whether problems encountered in the use of a method derive from the limitations of the method itself or from the inadequate resources of the practitioner.

In addition to collecting information about the assumptions embedded in, and the application of, the method being evaluated, it is most important to collect data on its *effects* as seen from the perspectives of those exposed to it (usually participants in workshops, but others might be relevant too, depending on the context). This is the crux of the evaluation of methods.

When considering effects, it is useful to distinguish between *process* and *outcome* criteria for their evaluation (e.g., Chess and Purcell, 1999; Rowe and Frewer, 2004): outcome criteria refer to whether, in a particular case, the method facilitated the achievement of specific goals (e.g., the production of a plan or the generation of a consensual vision), while process criteria refer to the effectiveness of the means by which these goals were achieved (e.g., did the method give everyone a chance to speak, allow creative exploration or enable a fair evaluation of options?). The difference between process and outcome can get a little blurred when an explicit goal of an intervention is, for instance, to facilitate participatory engagement. Nevertheless, keeping the distinction explicit helps us avoid potentially major mistakes like focusing so much on process that we fail to notice that people's purposes for the intervention have not been achieved, or focusing so much on outcomes that we miss negative effects of the process on participants.

Outcomes may also be longer term in nature, and these are not always predictable or easy to measure (Duignan and Casswell, 1989). Indeed, making a causal link between an intervention and an outcome that emerges, say, ten years later is often extremely difficult. Long-term follow up studies may be needed if some kinds of outcomes (e.g., those concerned with sustainability) are to be properly assessed, and sometimes it's simply a matter of maintaining contact with key individuals.

The usual means of measuring many process and short-term outcome effects of a method (other than through personal reflections by the researcher) is by gathering feedback from participants following workshops, often giving them questionnaires to fill in as soon as the workshop is complete (e.g., Duram and Brown, 1998; Rowe et al, 2004; Berry et al, 2006; Sykes and Goodwin, 2007). This is an approach that we have found valuable in our own systemic intervention practice, and we have developed a questionnaire with sections that are changeable from intervention to intervention to reflect specific local needs. Other sections are relatively stable and are used repeatedly across a variety of local intervention contexts. Both types of section are useful for locally meaningful evaluations, but the latter (stable) sections can also yield data for use in longer-term cross-method comparisons. More information about our questionnaire is provided below.

DEVELOPING AN EVALUATION QUESTIONNAIRE

Our questionnaire is not the only tool needed for evaluating systemic and participative methods (for instance, it cannot capture data on longer term outcomes). Nevertheless, it can make a useful contribution by gathering the viewpoints of participants immediately after their involvement in a workshop. The questionnaire has the following sections:

1. A five-point scale for the quantitative assessment of usefulness, plus open questions about what people liked and disliked, and what could have been done differently. Additional open questions reflecting local contingencies can be added in here if and when required.
2. Fifteen questions with five-point scales enabling the quantitative evaluation of whether certain things have been achieved. Both process and short term outcome questions are included here, and this is a set of questions that is not tailored to particular interventions (except occasional words where it is necessary to mention that the workshop is focused on a water, housing, health, policing, etc., issue). The process we went through to derive this set of questions is discussed below.
3. Thirteen questions, again with five-point scales, addressing potential negative attributes of (or things that can go wrong when using) systemic and participative methods. Once again this is an unchanging set of questions, and our process for deriving them is discussed below.
4. A set of open ended questions asking people to assess the process from their own cultural viewpoints. These questions are usually worded generally so they are relevant to multiple cultural perspectives, but specific questions relating to particular cultures can be added if required (for example, in New Zealand there often needs to be a specific focus on M_ori perspectives).
5. Questions gathering basic demographic data (sex, age, ethnic origin, etc.).

The Development Process

Our questionnaire was first developed in the context of a research programme aiming to generate and evaluate new participative and systemic methods for use in promoting sustainable resource use. This is why the focus is on *participative* methods and not others. We are looking to widen the boundary of our evaluative focus in the next year in the context of another project focused on modelling for drugs policy making.

The adaptable parts of the questionnaire (sections one, four and five above) were relatively straight forward to design, although they required some iterative testing to get them right. The more difficult task was to produce sections two and three, which needed to yield data for meaningful use in both local evaluations and longer term comparisons between methods. Because of the latter, the questions had to be reasonably generic. Other authors suggest a number of different ways of producing generic evaluation criteria, and these have been summarised by Beirerle and Konisky (2000) and Rowe and Frewer (2004). A combination of their thinking (plus an addition of our own) suggests that there are six distinct approaches: author-generated

(resulting from personal experience); practice-based (deriving from explicit reflections on case studies); theory-based (evaluating according to the expectations one would have if one agreed with a particular theory); literature-based (deriving from a review of other authors' work); expert-based (drawing on the views of an advisory panel); and survey-based (finding out from potential participants, either through interviews or a mail survey, what their most widely held expectations are). Some authors have combined two or more of the above.

Our own approach started with asking a key question: what do we want to measure? One option was to focus only on criteria that one would expect to be meaningful for all systemic and participative methods. This is the approach taken by Bjärås et al (1991) and Beirerle and Konisky (2000). However, while it is useful to identify 'common denominators' and assess methods against these, this does not help in evaluating the unique attributes of methods that might make them complementary rather than competing. To evaluate these, it is important to look at the set of possible common and divergent attributes that a range of systemic and participative methods might exhibit.

We therefore set out to identify a number of methodologies and methods that could fairly represent the diversity of participative systems approaches. We established a panel of six internationally known writers on systemic problem structuring, all of whom suggested candidate methods. We ended up with six participative systems approaches that all claimed to do different things. We then reviewed the literature on these, drawing out a set of attributes that could form the basis for questions to be asked of participants in workshops. We also asked the international panel to suggest their own evaluation criteria, and we added in a couple that were not apparent from the literature review but, in our experience, were important. This list was then sent back to the panel for peer review, resulting in some amendments. We ended up with a set of questions for field testing.

It is generally accepted that a questionnaire to be employed in an experimental context should be tested for validity (does it measure what we think it does?) and reliability (does it give consistent results?). However, for an evaluation questionnaire to be employed in the field outside the context of experimental studies, *usability* is just as important, if not more so (Rowe and Frewer, 2004). Usability means asking whether people are actually prepared to complete the questionnaire and do so in a sensible manner. Rowe and Frewer (2004) note that, because compromises have to be made in questionnaire design to ensure usability (e.g., the questions need to be answerable in five to ten minutes at the end of a gruelling day), usability is often inversely related to validity and reliability (both of which are enhanced by the generation of more rather than less data). This may be the case but, as Rowe et al (2005) say, there is no point even beginning to consider validity and reliability if the instrument cannot be used in the first place.

To check for usability, we field tested the questionnaire in five different interventions, each time making small amendments in response to issues thrown up by the way people approached the questions:

- Facilitating consultation with land owners and community interest groups as part of a feasibility study for the construction of a new water storage dam (Winstanley et al, 2005).
- Working with an Australian NGO and its stakeholders in exploring policy options to address the public injecting of illicit drugs (Midgley et al, 2005);
- Facilitating workshops with the police and other stakeholders in the criminal justice system to look at ethical issues associated with anticipated future developments of forensic DNA technologies (Baker et al, 2006);
- Reviewing the process used by the New Zealand Ministry of Research, Science and Technology to develop 'roadmaps' for long-term investments in environment, energy, biotechnology and nanotechnology research (Baker and Midgley, 2007); and
- Developing a new collaborative evaluation approach in partnership with regional council staff responsible for facilitating community engagement in sustainability initiatives (Hepi et al, 2007).

We also tested the questionnaire on interventions undertaken by people other than ourselves: a public meeting and a stakeholder forum convened in two different areas of New Zealand to discuss water shortages.

Following observations of participants completing the first version of the questionnaire, it was judged to be over-long. We shortened it, but then in later iterations found that there was a need to add in new questions. We ended up finding a compromise between comprehensiveness and brevity. On our first iteration of field testing, we also undertook a basic analysis to check that there were no counter-intuitive answers (which might suggest the misinterpretation of a question); that there was no tendency for people to tick the same point on all the scales (indicating boredom or a lack of comprehension); and that similar questions generated similar answers. All these checks proved satisfactory. Having undertaken this series of field tests, we are now reasonably confident of the usability of our questionnaire.

Interpreting Data Generated through Use of the Questionnaire

Before closing this discussion of our questionnaire, it is important to note that the data generated through it within the context of a particular systemic intervention always has to be interpreted in relation to the other aspects of our framework (context, purposes and practitioner skills and preferences). Failure to undertake analyses of these aspects could result in attributions to the method of results that might have had other origins.

However, in making longer term comparisons of methods using data from multiple case studies, we make the assumption that the more case studies are included, the more likely it is that the effects of particular contexts, purposes, etc., will be evened out.

STRENGTHS AND LIMITATIONS OF THE EVALUATION FRAMEWORK

As we see it, this new framework for the evaluation of methods has two significant strengths. First, by encouraging the exploration of the context-purposes-method relationship in a particular intervention, and by explicitly recognising that the practitioner becomes part of the situation that s/he intervenes in, our framework offers a more nuanced (but still reasonably parsimonious) set of concepts and guidelines to work with than many others in the literature. Second, it incorporates a questionnaire that can support both locally meaningful evaluations *and* longer-term comparisons between methods, thereby giving us the potential to move beyond the either/or debate that has characterised the literature in recent years.

Nevertheless, it is important to clarify some of the framework's limitations. In our view, the first two of these are more or less inevitable, and have to be managed as part of the evaluation process, while the final four indicate the need for further research. Only the first limitation concerns our framework as a whole: the rest relate solely to the use of the questionnaire for longer-term comparisons between methods.

Within the context of a specific use of a method in a single intervention, there is scope for the practitioner to avoid unwelcome conclusions, for example by exaggerating the effect of an aspect of context that was outside his or her control, thereby missing shortfalls in either the method or his or her own skill set. To help manage this, three methodological devices have been built into our framework to bring evidence of bad news to the attention of evaluators, making avoidance more difficult than it might be if the evaluators were basing their conclusions on personal reflections alone. First, the use of a questionnaire ensures that participant voices are available. In particular, the answers to the open ended questions are likely to include the participants' own theories about shortcomings. Second, by offering guidelines for exploring the context that draw upon multiple paradigmatic perspectives, the risk of 'paradigm blindness' (interpreting the context in the same paradigmatic terms as the method, thereby missing insights that would be apparent from other perspectives) is minimised (also see Romm, 1996). Third, by explicitly focusing attention on the practitioner's identity, purposes, skills and preferences, the framework confronts evaluators with some of the questions that they are most likely to want to avoid. If desired, and if feasible, practitioners can go one step further to minimise avoidance by including participants on the evaluation team (preferably ones that are themselves open to the possibility of receiving bad news).

The second limitation we are aware of, applying to longer term comparisons of methods using the questionnaire, comes from the observation that there is a strong movement advocating methodological pluralism or 'multi-methodology' (e.g., Flood and Jackson, 1991; Jackson, 1991; Flood and Romm, 1996; Mingers and Gill, 1997; Midgley, 2000). At its most flexible, a pluralist practice may involve the integration of several previously distinct methods into a new whole, perhaps also incorporating the design of novel elements (Midgley, 2000). It will be much easier to compare standard sets of methods (e.g., those associated with discrete systems methodologies) than it will be to compare hybrid approaches that have not been widely applied. The irony here is that the more flexible and responsive that systems practice becomes, the more

difficult it will be to evaluate methods over the longer term in a manner that can control for contextual effects. We certainly would not want to see our desire for improved evaluations of methods to result in the stultification of pluralist practice. Rather, we suggest that it may be wiser to accept that this limitation will restrict what can be asked of longer term comparisons between methods, but it will not make them redundant. It will still be possible to compare the sets of methods associated with well known and widely applied methodologies, giving us evidence of their strengths and weaknesses in relation to the set of attributes that a representative range of methods possesses. It will also be possible to compare pluralist practice in general with the use of particular discrete approaches. Finally, some hybrid approaches, if applied in several applications (we suggest ten as a minimum), can also be compared with other sets of methods. There are a number of relatively popular hybrids in the literature that will no doubt qualify for evaluation. When comparisons between hybrids using the questionnaire data look like they will be unreliable because the sample size is too small, it should nevertheless still be possible to facilitate cross-case study learning, where possible bringing together two or more research teams to reflect on their practice using our framework (Figure 1).

The third limitation is that we have not yet tested the questionnaire for validity and reliability. Rowe et al (2005) discuss the substantial difficulties in doing this in the field because participants are often reluctant to fill in two or more questionnaires asking similar things (the usual approach to testing for validity being to compare with another questionnaire constructed for similar purposes). Indeed, in this case, testing for validity will be difficult because there are only a couple of published instruments (e.g., Halvorsen, 2001), and they are geared to evaluating forms of public participation other than systemic problem structuring methods. Also, checking reliability is even more troublesome than a validity test because it involves getting participants to fill in the same questionnaire on two separate occasions. Generally speaking, the practitioner only has access to participants on the day of a workshop. Our intention is to do some validity and reliability testing in due course when a good comparative instrument can be identified and the testing can be added to an intervention without difficulty.

The fourth limitation we have identified concerns the inability of standard metrics, such as those to be found in sections two and three of our questionnaire, to pick up novelty: they can only evaluate against already established criteria. This is arguably one of the most significant limitations in terms of conducting longer-term research based on multiple case studies: it appears that, after around twenty years of relative stability in the number of systemic and participative problem structuring methods that are widely used in practice, a new generation of practitioners is now producing new methodologies and methods (Shaw et al, 2006; Franco et al, 2007), and it is important that the questionnaire does not go out of date. Our solution to this problem, which will need to be enacted as part of a longer term international research program, will be to undertake a review of the questionnaire after a set period of data collection. This period will need to be long enough to allow sufficient data to be gathered on the application of well established approaches. Periodic reviews of the questionnaire followed by new data collection should enable a balance to be struck between stability

(to facilitate robust comparisons) and change (to keep the longer term comparisons open to novelty).

The fifth limitation is that our questionnaire does not currently allow the comparison of participative and non-participative methods. Although our intention is to extend our research to include the latter, it may not be feasible to integrate questions about both participative and non-participative approaches into a single instrument. Our field testing suggests that we have already hit the upper limit for the number of questions people are willing to answer, so feasibility would depend on reducing the number of questions about participative methods in order to allow others to be included.

The sixth and final limitation we face is that no one group of practitioners will be able to collect sufficient data on its own to enable the robust, longer term comparison of methods. International collaboration will therefore be essential, as will a central resource for data collection and analysis. This means securing funding for the central resource; seeking agreement between collaborators on the data to be collected; and ensuring commitment to use of the agreed instrument. We believe that we have made a reasonable start in moving towards an international collaboration by developing a questionnaire that seeks to identify the complementary strengths of different methods rather than evaluating them against a small number of common denominators (which would set up a competitive environment). Nevertheless, it will probably be wise to once again review our instrument at the start of new research to ensure collaborators' confidence in it.

CONCLUSION

In this paper we have offered a new framework for evaluating systemic and participative methods, focusing on the context-purposes-method relationship. This framework can be used in an action research mode, and it asks practitioners to view themselves as active contributors to the success or failure of a method-in-context. We have also reported on the development of a questionnaire to gather data from participants that can be of use in reflecting on the strengths and weaknesses of methods. The same data may be useful for both evaluations of methods in single case studies *and* longer term comparisons between methods using information from multiple cases. However, undertaking longer term comparisons will require a new, international research program. We would like to end by calling for collaborators. If we can gather together an international group of systems practitioners who want to enhance the evaluation of methods through both single case studies and longer term cross-method comparisons, then we will be in a good position to apply for funding for a central resource to co-ordinate our efforts. Please contact Gerald Midgley < gerald.midgley@esr.cri.nz > to discuss possibilities.

ACKNOWLEDGEMENTS

We would like to acknowledge the support of colleagues at ESR who helped to field test the questionnaire in the context of their projects: Virginia Baker, Jan Gregor, Wendy Gregory, Maria Hepi, Miria Lange, Johanna Veth and Ann Winstanley. We would also like to recognise the contribution made by our international panel of

experts on systemic problem structuring methods: John Brocklesby, José Córdoba, Amanda Gregory, John Mingers, Leroy White and Jennifer Wilby. Finally, we need to acknowledge funding from various sources that enabled the research reported in this paper: the Foundation for Research, Science and Technology (contracts C03X0304 and C03X0305); the Colonial Foundation Trust; and the New Zealand Ministry for Research, Science and Technology.

REFERENCES

Adams, R., and McCullough, A. (2003). The Urban Practitioner and Participation in Research within a Streetwork Context, *Community, Work & Family*., 6(3):269-287.

Alberts, D.J. (2007). Stakeholders or Subject Matter Experts, Who should be Consulted? *Energy Policy*., 35:2336-2346.

Allsop, J., and Taket, A. (2003). Evaluating User Involvement in Primary Healthcare, *International Journal of Healthcare Technology & Management*., 5(1/2):34-44.

Alrøe, H.F. (2000). Science as Systems Learning: Some Reflections on the Cognitive and Communicational Aspects of Science, *Cybernetics and Human Knowing*., 7:57-78.

Baker, V., Gregory, W., Midgley, G., and Veth, J. (2006). *Ethical Implications and Social Impacts of Forensic DNA Technologies and Applications: Summary Report*., ESR, Christchurch.

Baker, V., and Midgley, G. (2007). *Review of the MoRST Roadmaps Exercise: Final Report*., Confidential ESR Client Report, ESR, Wellington.

Bateson, G. (1970). Form, Substance, and Difference, *General Semantics Bulletin*., 37:5-13.

Beierle, T.C., and Cayford, J. (2002). *Democracy in Practice: Public Participation in Environmental Decisions*., RFF Press, Washington DC.

Beierle, T.C., and Konisky, D.M. (2000). Values, Conflict, and Trust in Participatory Environmental Planning, *Journal of Policy Analysis and Management*., 19(4):587-602.

Berry, H., Bowman, S.R., Hernandez, R., and Pratt, C. (2006). Evaluation Tool for Community Development Coalitions, *Journal of Extension*., 44(6): http://www.joe.org/joe/2006december/tt2.shtml (accessed: 30.3.07).

Bjärås, G., Haglund, B.J.A., and Rifkin, S.B. (1991). A New Approach to Community Participation Assessment, *Health Promotion International*., 6(3):199-206.

Branch, K.M., and Bradbury, J.A. (2006). Comparison of DOE and Army Advisory Boards: Application of a Conceptual Framework for Evaluating Public Participation in Environmental Risk Decision Making, *Policy Studies Journal*., 34(4):723-753.

Buysse, V., Wesley, P., and Skinner, D. (1999). Community Development Approaches for Early Intervention, *Topics in Early Childhood Special Education*., 19(4):236-243.

Charnley, S., and Engelbert, B. (2005). Evaluating Public Participation in Environmental Decision-Making: EPA's Superfund Community Involvement Program, *Journal of Environmental Management*., 77:165-182.

Checkland, P. (1981). *Systems Thinking, Systems Practice*., Wiley, Chichester.

Checkland, P., and Scholes, J. (1990). *Soft Systems Methodology in Action*., Wiley, Chichester.

Chess, C., and Purcell, K. (1999). Public Participation and the Environment: Do We Know what Works? *Environmental Science & Technology*., 33(16):2685-2692.

Churchman, C.W. (1970). Operations Research as a Profession. *Management Science*., 17:B37-53.

Clayton, A.M.H., and Radcliffe, N.J. (1996). *Sustainability: A Systems Approach*., Earthscan, London.

Cole, M. (2006). Evaluating the Impact of Community Appraisals: Some Lessons from South-West England, *Policy & Politics*., 34(1):51-68.

Connell, N.A.D. (2001). Evaluating Soft OR: Some Reflections on an Apparently 'Unsuccessful' Implementation Using a Soft Systems Methodology (SSM) Based Approach, *Journal of the Operational Research Society*., 52:150-160.

Douglas, M. (1986). *How Institutions Think*., Routledge and Kegan Paul, London.

Duignan, P., and Casswell, S. (1989). Evaluating Community Development Programs for Health Promotion: Problems Illustrated by a New Zealand Example, *Community Health Studies*., 13(1):74-81.

Duram, L.A., and Brown, K.G. (1999). Assessing Public Participation in U.S. Watershed Planning Initiatives, *Society & Natural Resources*., 12:455-467.

Eden, C. (1995). On Evaluating the Performance of 'Wide-Band' GDSS, *European Journal of Operational Research*., 81:302-311.

Entwistle, V., Buchan, H., Coulter, A., and Jadad, A. (1999). Towards Constructive Innovation and Rigorous Evaluation: A New Series on Methods for Promoting and Evaluating Participation, *Health Expectations*., 2:75-77.

Flood, R.L. (1995). *Solving Problem Solving*., Wiley, Chichester.

Flood, R.L., and Jackson, M.C. (eds.) (1991). *Critical Systems Thinking: Directed Readings*., Wiley, Chichester.

Flood, R.L., and Romm, N.R.A. (eds.) (1996). *Critical Systems Thinking: Current Research and Practice*., Plenum, New York.

Forrester, J.W. (1969). *Principles of Systems*., Wright-Allen Press, Cambridge MA.

Franco, A., Shaw, D., and Westcombe, M. (2007). Taking Problem Structuring Methods Forward, *Journal of the Operational Research Society*., 58(5):545-546.

Halvorsen, K.E. (2001). Assessing Public Participation Techniques for Comfort, Convenience, Satisfaction, and Deliberation, *Environmental Management*., 28(2):179-186.

Hepi, M., Ahuriri-Driscoll, A., Foote, J., Midgley, G., and Winstanley, A. (2007). Research report currently in preparation., ESR, Christchurch.

Ho, C.H. (1997). *A Critical Process for the Evaluation of Methodology*., Ph.D. thesis, University of Hull.

Jackson, M.C. (1991). *Systems Methodology for the Management Sciences*., Plenum, New York.

Jackson, M.C., and Keys, P. (1984). Towards a System of Systems Methodologies, *Journal of the Operational Research Society*., 35:473-486.

Jenkins, N.T., and Bennett, M.I.J. (1999). Toward an Empowerment Zone Evaluation, *Economic Development Quarterly*., 13(1):23-28.

Kelly, K., and Van Vlaenderen, H. (1995). Evaluating Participation Processes in Community Development, *Evaluation & Program Planning*., 18(4):371-383.

Li, X., and Zheng, H. (1995). Study on General Systems Methodology, in *Systems Methodology: Possibilities for Cross-Cultural Learning and Integration*, (G. Midgley, and J. Wilby, eds.), Centre for Systems Studies, Hull.

Luhmann, N. (1986). *Ecological Communication.*, University of Chicago Press, Chicago.

Margerum, R.D. (2002). Collaborative Planning: Building Consensus and Building a Distinct Model for Practice, *Journal of Planning Education & Research*., 21:237-253.

Masozera, M.K., Alavalapati, J.R.R., Jacobson, S.K., and Shrestha, R.K. (2006). Assessing the Suitability of Community-Based Management for the Nyungwe Forest Reserve, Rwanda, *Forest Policy & Economics*., 8:206-216.

McAllister, K. (1999). Understanding Participation: Monitoring and Evaluating Process, Outputs and Outcomes, *Working Paper 2*., IDRC, Ottawa.

McGurk, B., Sinclair, A.J., and Diduck, A. (2006). An Assessment of Stakeholder Advisory Committees in Forest Management: Case Studies from Manitoba, Canada, *Society & Natural Resources*., 19:809-826.

Midgley, G. (1994). Ecology and the Poverty of Humanism: A Critical Systems Perspective, *Systems Research*., 11:67-76.

Midgley, G. (2000). Systemic *Intervention: Philosophy, Methodology, and Practice*., Kluwer/Plenum, New York.

Midgley, G., Ahuriri-Driscoll, A., Baker, V., Foote, J., Hepi, M., Taimona, H., Rogers-Koroheke, M., Gregor, J., Gregory, W., Lange, M., Veth, J., Winstanley, A. and Wood, D. (2007). Practitioner Identity in Systemic Intervention: Reflections on the Promotion of Environmental Health through M_ori Community Development, *Systems Research and Behavioral Science*., 24:233-247.

Midgley, G., Winstanley, A., Gregory, W., and Foote, J. (2005). *Scoping the Potential Uses of Systems Thinking in Developing Policy on Illicit Drugs*., Drug Policy Modelling Project Research Memorandum #13, Turning Point, Melbourne.

Mingers, J.C. (1997). Towards Critical Pluralism, in *Multimethodology: The Theory and Practice of Combining Management Science Methodologies*, (J. Mingers and A. Gill, eds.), Wiley, Chichester.

Mingers, J., and Gill, A. (eds.) (1997). *Multimethodology: The Theory and Practice of Combining Management Science Methodologies*., Wiley, Chichester.

Mingers, J., and Rosenhead, J. (2004). Problem Structuring Methods in Action, *European Journal of Operational Research*., 152:530-554.

Morgan, L.M. (2001). Community Participation in Health: Perpetual Allure, Persistent Challenge, *Health Policy & Planning*., 16(3):221-230.

Murphy-Berman, V., Schnoes, C., and Chambers, J.M. (2000). An Early Stage Evaluation Model for Assessing the Effectiveness of Comprehensive Community Initiatives: Three Case Studies in Nebraska, *Evaluation & Program Planning*., 23:157-163.

Ong, B.N. (2000). Assessing the Context for Partnerships between Communities and the National Health Service in England, *Critical Public Health*., 10(3):343-351.

Romm, N.R.A. (1996). Inquiry-and-Intervention in Systems Planning: Probing Methodological Rationalities, *World Futures*., 47:25-36.

Rowe, G., and Frewer, L.J. (2000). Public Participation Methods: A Framework for Evaluation, *Science, Technology & Human Values*., 25(1):3-29.

Rowe, G., and Frewer, L.J. (2004). Evaluating Public Participation Exercises: A Research Agenda, *Science, Technology & Human Values*., 29(4):512-556.

Rowe, G., Horlick-Jones, T., Walls, J., and Pidgeon, N. (2005). Difficulties in Evaluating Public Engagement Initiatives: Reflections on an Evaluation of the UK *GM Nation?* Public Debate about Transgenic Crops*, Public Understanding of Science*., 14:331-352.

Rowe, G., Marsh, R., and Frewer, L.J. (2004). Evaluation of a Deliberative Conference, *Science, Technology & Human Values*., 29(1):88-121.

Shaw, D., Franco, A., and Westcombe, M. (2006). Problem Structuring Methods: New Directions in a Problematic World, *Journal of the Operational Research Society*., 57(7):757-758.

Shaw, I. (1999). *Qualitative Evaluation*., Sage, London.

Shen, C-Y., and Midgley, G. (2007). Toward a Buddhist Systems Methodology 1: Comparisons between Buddhism and Systems Theory. *Systemic Practice and Action Research*., 20(3):167-194.

Sieber, R. (2006). Public Participation Geographic Information Systems: A Literature Review and Framework, *Annals of the Association of American Geographers*., 96(3):491-507.

Smith, L.T. (ed.) (1999). *Decolonizing Methodologies: Research and Indigenous Peoples*, Zed Books, London.

Sykes, C., and Goodwin, W. (2007). Assessing Patient, Carer and Public Involvement in Health Care, *Quality in Primary Care*., 15:45-52.

Tuler, S., Webler, T., and Finson, R. (2005). Competing Perspectives on Public Involvement: Planning for Risk Characterization and Risk Communication about Radiological Contamination from a National Laboratory, *Health, Risk & Society*., 7(3):247-266.

Ulrich, W. (1983). *Critical Heuristics of Social Planning: A New Approach to Practical Philosophy*., Haupt, Berne.

Warburton, D., Wilson, R., and Rainbow, E. (2007). *Making a Difference: A Guide to Evaluating Public Participation in Central Government*., Involve, London, http://www.involve.org.uk/evaluation (accessed: 30.5.07).

White, L. (2006). Evaluating Problem-Structuring Methods: Developing an Approach to Show the Value and Effectiveness of PSMs, *Journal of the Operational Research Society*., 57:842-855.

Winstanley, A., Baker, V., Foote, J., Gregor, J., Gregory, W., Hepi, M., and Midgley, G. (2005). *Water in the Waimea Basin: Community Values and Water Management Options*., ESR, Christchurch.

Yearley, S. (2006). Bridging the Science-Policy Divide in Urban Air-Quality Management: Evaluating Ways to Make Models More Robust through Public Engagement, *Environment and Planning C*., 24:701-714.

Zhu, Z. (2000). Dealing with a Differentiated Whole: The Philosophy of the WSR Approach. *Systemic Practice and Action Research*., 13(1): 21-57.