Author(s): Sampson, Alice.
Article Title: Developing Robust Approaches to Evaluating Social Programmes
Year of publication: 2007
Citation: Sampson, A. (2007) 'Developing Robust Approaches to Evaluating Social Programmes' Evaluation, 13 (3) 477-493
Link to published version:
http://dx.doi.org/10.1177/1356389007082132
DOI: 10.1177/1356389007082132

Publisher statement:
http://www.uk.sagepub.com/authors/journal/permissions.sp

Developing robust approaches to evaluating social programmes

Alice Sampson

Evaluation The International Journal of Theory, Research and Practice 2007 13(3): 477-493

Abstract

Using 'evidence' to falsify rather than verify patterns in data and searching for alternative explanations, enables a better understanding of the circumstances which explain why and how a social programme works or does not work. An analysis of the extent to which a programme is meeting its aims and objectives, and the policy itself to find out if it provides a solution to the policy problem, is more rigorous. The roles researchers adopt influences the quality of an evaluation; facilitating a better understanding of the theories embodied in programmes enhances an evaluation whilst being a 'broker of compromise' can limit access to information. Researchers have a valuable role in promoting learning. A robust evaluation framework integrates strategies for generalising at the outset and identifying mechanisms of change or causal mechanism is a way forward. Examples are taken from recent evaluations conducted by the author and colleagues to illustrate the arguments.

Keywords/key phrases

Theory driven evaluation, use of evidence, role of researchers, generalisability.

Author Biography

Alice Sampson leads the urban regeneration research team at the Centre for Institutional Studies, University of East London, and is involved in evaluating the impact of recent government initiatives to alleviate poverty and disadvantage in local areas.

Introduction

In recent years there has been an increased commitment by central government to commissioning evaluations of its new social policies. This has led to the development of theoretical approaches to evaluation, most notably in Britain, Realistic Evaluation (Pawson and Tilley 1997) and in America, Chen (1990), Weiss (1997) and the Theories of Change (ToC) approach by the Aspen Institute (Connell et al 1995;

Kurbisch et al, 1999). The intention of these approaches is to develop robust approaches to evaluating social programmes using theories to explain changes that have occurred due to the interventions.

In this paper developing robust approaches to evaluating social programmes is considered from a belief that the purpose of evaluating social programmes is to contribute to social improvement. From this perspective a robust approach integrates a range of factors into an evaluation framework including issues about how an evaluation is conducted and how the findings are used. Four issues considered integral to a robust evaluation framework are discussed in this paper: the use of 'evidence', the evaluation of a programme's aims and the persistence of the problem the policy is designed to alleviate, the role of researchers in contributing to learning, and the transfer of successful programmes or certain aspects of a programme to other locations.

To set the scene the problems typically faced by evaluators are outlined to explain why a theoretical perspective to evaluation is preferred.

Understanding the research problem

The challenges faced by evaluators have been well described by Kubisch and colleagues (1995). They recognise the complexities of social programmes which aim to address multi-faceted problems and seek solutions that can address a multitude of problems including crime, housing, health and physical infrastructure, with intended improvements aimed at an individual, family, community and institutional level. They also recognise that programmes evolve, are responsive to changing circumstances and contextual issues such as political and financial systems. The problem for evaluators is to identify all the changes that are taking place, to measure them, and to assess if the changes are due to the social programme, to other extraneous factors, or would have happened anyway (Connell et al, 1995; Kubisch et al, 1995; Pawson and Tilley, 1997).

A theoretical approach assists researchers address these complexities; it encourages researchers to examine the assumptions underlying the programme and makes explicit the link between activities and outcomes (Connell et al, 1995; Pawson and Tilley, 1997). By developing a causal chain of explanations, the reasons why a policy initiative works, and how it works, can be established (Weiss, 1995). The approach also benefits practitioners. It gives them the opportunity to think about the links between the aims and objectives of an initiative and activities they intend to put in place or already have in place.

For these reasons we[i] are sympathetic to a theoretical base for evaluations and concur that the main focus of an evaluation is to conduct an impact study (Weiss, 1972). From this starting point we examine the different ways in which 'evidence' is used by researchers.

The use of 'evidence'

In this section an essentially Popperian argument is made that using 'evidence' to falsify rather than verify patterns in data, and searching for alternative explanations and anomalies leads to a more robust approach to evaluation. It is also argued programmes which have been successful in the past will not necessarily be successful in the future.

Typically researchers use information and data they have collected to either verify or falsify the theory that explains the impact of the social programme. An inductive approach is clearly articulated by Connell and Kurbisch (1999) when they state that:

> 'The theory of change approach contends that the more the events predicted by theory actually occur over the course of the CCI [Complex Community Initiative], the more confidence evaluators and others should have that the initiative's theory is right' (Connell and Kubisch, 1999: 9)

The idea that accumulative evidence gives evaluators more confidence in the robustness of their findings has however, been highlighted by Popper as problematical:

> 'No matter how many instances of white swans we may have observed, this does not justify the conclusion that all swans are white' (Popper, 1968:27. Italics in the original).

Logically, as Popper points out, just one exception can refute a general theory whilst no number of confirming instances can establish it is correct. Furthermore an examination of the exceptions or anomalies rather than the consistencies can provide more insights into, and explanations of, the general pattern. A search for the 'the black swan' enables researchers to find out about different circumstances which explain why the swan is black instead of the anticipated white and this enables us to understand more about the particular circumstances which explain the existence of the white swans as the dominant type. Thus finding out about the circumstances that explain why a project has not worked can inform policy-makers and practitioners about the conditions for successful initiatives.

Evidence is often used as an indication that a programme can be 'rolled out' across a wide range of locations. However Popper argued that, in fact, theory is conjecture; what has happened in the past will not necessarily occur in the future and the conditions and circumstances under which there was a cause and effect in the past will not necessarily continue to happen in the future (Popper, 1968). The classic example is that for many years people believed that the world was flat, until new knowledge disproved this belief.ii Similarly as evaluators we collect and collate information on past trends, events, experiences and practices, and apply theories developed from this retrospective information to present circumstances. We cannot be sure that the explanations arising from this research will hold; with changed circumstances and new knowledge the causal mechanisms may no longer be active in the new situations (Sayer 1992; Pawson and Tilley 1997). As researchers we cannot be certain that the 'evidence' will predict what will happen in the future.

The Popperian approach of falsification requires a search for alternative explanations using data to 'test' each hypothesis until hypotheses are chosen where the data is accommodated. As evaluators we try to provide the best explanation. In this way data are used to establish 'beyond reasonable doubt'iii that a particular explanation is the most appropriateiv. Importantly we also 'test' hypothesis to find out about any harm a social programme may be causing.

The challenge for researchers is to decide when to stop formulating alternative hypotheses. To be rigorous an evaluator disproves as many hypotheses as possible to find alternative explanations because it is through this process that researchers can make more informed assessments about the most plausible effect of a programme. Often purely practical constraints such as the size of the research grant and limited time affect the thoroughness of the evaluation (Weiss, 1997; Griffith and Sampson, 2002). Even if practical constraints are minimised, there will always remain uncertainties about what is happening in the present and what will happen in the future. A good evaluation will reduce these uncertainties by clarifying the benefits and disadvantages of a programme to enable policy-makers and practitioners to make better informed decisions and to enable them to understand the implications of these decisions (Weiss 1972:4). But uncertainties will remain and for this reason any claim that theories are 'right' does not reflect what evaluations can honestly achieve; we can never be certain that a theory is right.

The following example of an evaluation is designed to illustrate how using accumulative evidence can be misleading and that by refuting assumptions contained within social programmes explanations about the effectiveness of a programme are improved. We undertook an evaluation of three advocacy projects which all aimed to reduce incidents of domestic violence against women (Parmar and Sampson, 2005a and 2005b). The projects had the same aim, to reduce domestic violence, and similar objectives, outputs and outcomes. They all employed advocates on the proposition that if women domestic violence survivors were supported by advocates then they would make the decision to take actions that would lead to the cessation of the violence. The intention of the projects was to enable women to use legal remedies to obtain justice for the crimes committed against them.

The accumulative evidence indicated that one project was performing more effectively than others. Where advocates worked intensively with women the violence was more likely to stop; in this project the advocates worked intensively with 22% of the women compared with 13% and 8% in the other two projects. From a cost effectiveness perspective the unit cost per referral actively worked with – that is some work was completed with women – was significantly lower in this project; £317 compared to £835 and £850 in the other two projects (Parmar, Sampson and Jordan, 2003).

We developed a series of hypotheses to disprove that advocacy was working and we regarded the 'successful' project as the anomaly or outlier. In our search for alternative explanations we made the judgement that the most plausible explanation for the 'success' was that referring agencies perceived that the main purpose of the project was to assist women take legal action against the perpetrators of the violence. For this reason, women who had already made up their mind to end their violent relationship and who were willing to seek legal remedies, were most frequently referred to the project. In these cases, advocates were not changing or influencing the decisions women made to leave a violent relationship. This process of informal selection was not in place in the two other projects. The women referred to these projects were not so intent on becoming involved in the project as a means of ending their violent relationship. The advocates in these projects found it harder to engage women referred to them and to influence the decisions made by women that may lead to a cessation of violence. Thus we could not conclude that the apparently successful project was indeed 'successful' because for many cases advocacy was not being fully 'tested' – the advocates did not have to influence the decision making of women to end the violence; they worked with women to facilitate the progress of a woman's case through legal the legal system.v

Evaluating aims and problem solving

Researchers are typically commissioned to evaluate the performance of a social programme by assessing the extent to which it has met its aims. The aims are usually measured by a set of agreed outcomes and the attention of researchers is therefore focused on the programme itself; how it is performing and deciding how to measure the outcomes to reflect the activities of the programme. The ToC framework provides an example of this approach; the focus is on the theories of change which occur within the programme - the programme logic, and the validity of these theories are assessed during an evaluation (Cornell and Kubisch, 1999).

This approach is clearly useful to policy-makers and practitioners enabling them to learn lessons from experimental social programmes. Such approaches to evaluation have, however, the potential to lose sight of the initial purpose of a social programme, which is essentially a policy response to an identified problem. We would argue that a rigorous evaluation includes an analysis of whether the programme is in fact an appropriate solution to the problem identified by policy makers (Burgess, 2002). In this way a rigorous evaluation is one which has a dual agenda; assessing the extent to which aims have been achieved and asking if the policy itself provides a solution to the problem it is designed to address.

The starting point for evaluating the 'appropriate policy' approach is the proposition that policies themselves are theories (Majone, 1980), are tentative solutions to a particular social problem (Burgess, 2002), and the hypotheses upon which they rest can be refuted. The value of this approach is illustrated by a recent evaluation we completed of a local initiative which was part of the Sure Start national programme designed to improve the physical, intellectual and social development of young children, particularly amongst those who are disadvantaged, as one solution to the problem of social exclusion. The intention of the programme is to improve the health of babies, to enrich their experiences of learning through quality play and childcare and provide support to families with respect to improving their knowledge about raising children, and by providing increased opportunities for employment to address the problem of workless households. The findings from the Sure Start programme presented below are those relevant to the research question: Is the Sure Start programme an appropriate solution to the problem of lack of opportunities for young children?

To select an area for the local Sure Start the local authority understandably used the Department of Transport, Local Government and the Regions Indices of Deprivation for 2000 (National Statistics Online, 2003) and chose the ward which was according to the index, unambiguously the most disadvantaged. In order to understand the problems of disadvantage in the area we analysed the components of the index and we interviewed, in their homes, a sample of 46 parents/carers with children under the age of 5 years and asked them about their experiences of raising a family in the area. The sample was not random as there were a high proportion of non-respondents so other easier to reach parents/carers were interviewed in their place. It is unclear exactly how this process of including different addresses in the sample once we had at least three no replies or a refusal to participate, affected the findings.

As shown in table 1 below the access domain ranking demonstrates that the ward is within the best 10 per cent for the country; it is located close to a large shopping centre with a range of services and has excellent public transport links.[vi] Health services were highly regarded with a satisfaction rate of over 92% for a range of pre and post natal services, and the majority of respondents claimed that they and their families enjoy good health which concurs with the health domain information collated for the deprivation indices. The biggest health problem was identified as emotional or mental problems (24%) and the main reasons given for this were poor housing, redundancy or financial problems. Although reducing smoking by parents is a target for the programme, only one mother admitted smoking, and four respondents had family members who smoked in front of the children. This compares to 55% in another Sure Start area, where the programme target is more appropriate (Brookes, 2002).

Only four respondents had no educational qualifications, most had General Certificates of Secondary Education (GCSEs) or equivalent, and almost a quarter (22%) had higher education qualifications. When researchers discussed parents/carers aspirations for their children, they recognised that the achievement of high educational standards was the route to a better quality of life, including high earnings.

Table 1. Comparison of East London Ward Ranking with Other English Wards

Indices of deprivation for East London ward, 2000	*East London ward rank compared to 8414 wards*	*East London ward rank compared to 8414 wards (%)*
Access domain	7632	90.70
Health domain	2274	27.00
Education domain	1227	14.60
Employment domain	1183	14.10
Child poverty index	1181	14.00
Income domain	830	9.90
Housing domain	62	0.70
Index of multiple deprivation rank	904	10.7

Source: www.neighbourhood.statistics.gov.uk (Department of Transport, Local Government and the Regions, Indices of Deprivation for wards in England, 2000)
Note: The ward ranked 1 in the indices of deprivation is the most disadvantaged ward in the country, and the ward ranked 8414 is at the top of the scale, and the most prosperous.

Table 1 shows that the income domain is within the worst 10% of all the wards included in the indices of deprivation index and one of the main problems for those raising a young family. The number of households with no one in employment was lower (24%) than other Sure Start areas; there were 44% in the Euston Sure Start (McQuail and Wigfall, 2001) and in Sure Start Abbey 48% were out of work (Cordis Bright, 2003), for example. We also found that majority of those 'not in employment' were not seeking work whilst their children were still babies, making it difficult to put in place measures to meet the Sure Start programme target of reducing workless households particularly where single parents chose not to work. The particularly difficult problem was low income, due to unskilled employment such security, construction or cleaning work and assistants at schools and of those in work a significant percentage had temporary employment (42%). Almost half the households received income support to supplement their earnings.

The indices of deprivation clearly identify the housing domain as the main problem in the area, a problem which is one of the worst in the country (see table 1). From the interviews we found out that the few blocks of flat on the edge of the area were in disrepair with central heating not working, draughts and families living in over-crowded conditions. A high proportion of the terraced houses was owned on a mortgage (54% of the respondents had a mortgage).

The above analysis of the problem of disadvantage in one particular Sure Start area demonstrates the value of an evaluation examining the nature of the disadvantage and its causes. Some of the nationally set targets of the programme such as reducing smoking amongst mothers and reducing the proportion of workless households do not address the main problems of poor housing, low wages and insecure employment found in this area. Poor maintenance and over-crowding were issues for those living in public housing and for homeowners low incomes affect the ability of householders to pay mortgages and maintain their homes. Until the main problems experienced by parents/carers raising young families in the area are addressed, it is unlikely that the Sure Start programme will offer the people living in this area any real routes out of poverty and disadvantage.

Role of researchers

This section illustrates how some roles taken on by researchers contribute to the rigour of the evaluation whilst others can detract from the quality of the research. It shows how the value of an evaluation can be strengthened where there is a well-developed theory underpinning a role taken on by researchers.

It is proposed that an evaluator has a core distinctive role; namely to assess the impact of a social programme by using their skills as a researcher to develop a robust evaluation framework; by adopting appropriate methods and applying them to a high standard; and by assessing the progress of the initiative by making judgements based on the research data and from the perspective of the values embodied in the evaluation framework (see for example Pawson and Tilley 1997; Oakley 2000).

To perform this role researchers typically require access to a number of databases, some of which may not be easily accessible and the goodwill of agencies will be required for the information to be released. Similarly when researchers are interviewing participants and beneficiaries of a social programme, the quality of the information obtained will be influenced by the relationship a researcher establishes with the interviewee. Thus any role adopted by researchers should avoid jeopardising access to, and the disclosure of, information. This position is often difficult for researchers to establish and maintain as the evaluation process typically takes place in a highly politicised environment within which the evaluation itself becomes politicised (Israel *et al*, 1992; Weiss, 1997; Marshall, 2001). Many problems can ensue including resistances to an evaluation which make it difficult for a researcher to perform their tasks.

The communication of research findings are also an innately political process as they include value judgements about 'what ought to be' and different 'ideas or meanings of what ought to be' which will lead to different assessments of the progress of the programme (Chen, 1990:57). The scope for giving different interpretations to findings means that findings can be used by stakeholders to further their own interests or to support particular interventions rather than others. Tensions can also arise between researchers and commissioning agents who want a 'spin' put on disappointing research findings (Fearnley and McInroy, 2000). Any differences or disagreements between evaluators and stakeholders have the potential of 'closing down' access to data and information which undermines the quality of the research, particularly where the evaluator looses credibility (Brown, 1995: 213).

Given the difficulties surrounding the core role of an evaluator, adopting additional roles should avoid making the situation even more difficult and, at the same time, should enhance the quality of the evaluation. All possible roles that evaluators could adopt are not included in the discussion but rather the intention is to indicate what factors may be considered when incorporating researchers' roles into an evaluation framework. By way of illustration an example is given below. It is argued that facilitating a better understanding of the processes of change may enhance the quality of an evaluation as this knowledge enables researchers to formulate additional hypotheses and informs them what data is most likely to sensitively measure changes, whilst acting as brokers of compromise between stakeholders is likely to lead to fewer opportunities to obtain data necessary to 'test' hypotheses.

Improving understandings of the processes of change

One possible role for evaluators is to improve practitioners' understandings of how their proposed interventions will work in practice. This approach is advocated by proponents of the ToC who argue that evaluators should be involved from the conception of an initiative to work with practitioners to surface and articulate why a social programme may be expected to work and to facilitate the identification of outcomes, the ways in which outcomes can be measured, and the collection of the relevant data (Connell and Kubisch, 1999). Through this work researchers are able to gain a better understanding of the intentions of the initiative. This will assist in improving the quality of the evaluation through an improved identification of the expected causal links between the activities and outcomes, and the improved identification of relevant information to measure anticipated impacts. It also has the potential to improve the intervention itself as the process of encouraging practitioners to articulate the theories of change embodied in the intervention can lead to a better understanding of how the interventions are likely to work and therefore what should be done to achieve social improvements (Pawson and Tilley, 1997; Connell and Kubisch, 1999, for example).

In adopting this role researchers act as facilitators to encourage practitioners to examine the assumptions contained within the programme they are intending to implement, and to articulate what changes might arise from these assumptions and how they might be measured in order to assess their impact. In this process the researcher's role is founded on research-based knowledge and retains their distance from the implementation of the programme; how practitioners use the knowledge they have gained from working with the researcher remains the responsibility and decision of the practitioners.

In some studies researchers have taken on an additional role and have become part of the implementation process itself. Where evaluators have this role they become associated as a member of the implementation 'team' and they are part of the decision-making about the future direction of the initiative and integral to delivering the interventions. For example, the ToC framework recognises that initiatives are likely to have multiple theories of change operating and that stakeholders may have competing theories of change. The researcher is assigned the role of facilitating stakeholders to compromise (Connell and Kurbisch, 1999). Some proponents of the approach claim that researchers have the skills to achieve this task and that they should adopt this role; 'skilled evaluators can gain consensus amongst the main parties involved in implementing community initiatives' (Judge and Bauld, 2001:25) and encourage stakeholders 'to compromise on some issues' (Macaskill et al, 2000:67 quoted in Judge and Bauld, 2001). Although researchers may have the skills to be effective brokers of compromise this role is not necessarily compatible with undertaking an effective outcome study for the following reasons.

Negotiating between different groups and organisations with the aim of improving the content of the programme gives researchers' responsibilities commensurate with being a stakeholder, and their accountability shifts from that of a researcher to an implementer. In their experiences of utilising the ToC framework Mackenzie and Blamey (2005) found that a blurring of roles occurred between evaluators and project staff, and a confusion ensued which lead to a host of misunderstandings as well as a lack of accountability.

Taking on the role of a broker of compromise adds further complexities to the task of an evaluator and if, during the process of brokering compromises, evaluators alienate or antagonise stakeholders, then it is highly likely that the research will be detrimentally affected. For example it is harder for researchers to establish trust with a stakeholder in an interview and this situation adversely affects the quality of the dialogue between the interviewer and interviewee. Even in situations where there are no obvious conflicts, any clear benefits with respect to obtaining better information to enhance the quality of the research are not always apparent (see for example Mackenzie and Blamey, 2005). Furthermore once researchers start to negotiate consensus and broker compromises their position becomes less clear and less transparent and therefore less open to scrutiny and criticism. [vii]

Promoters of learning

To facilitate a better understanding of the processes of change at a programme level researchers can adopt a role as promoters of learning during the research itself. The proposition is that where new learning takes place the opportunities for social improvement are greater and that researchers can have a role in creating these learning opportunities without jeopardising their role as evaluators; that is without limiting access to information, or undermining the quality of information they may obtain during interviews.

Considering how researchers may best promote learning in order to improve policies and practices has been highlighted as a significant but complex issue (Connell and Kubisch 1999; Nutley *et al*, 2003). For some the very purpose of doing an evaluation is precisely *'to ensure that lessons from experiments are learned in a systematic way so that they can be applied to the next generation of policies, programs, and research'* (Kubisch et al, 1995:16). But the difficulties involved in incorporating new knowledge gained from evaluations into policy making and into developing practices have been repeatedly highlighted (Kubisch *et al*, 1995; Lincoln, 2001; Nutley *et al*, 2003, for example). To meet these challenges a robust evaluation framework is one which gives researchers the opportunity to create the conditions necessary for learning to take place and for learning to be transformed into action. It is argued that this is more likely to occur where a relevant theory of learning is incorporated into the evaluation framework. Two approaches are discussed; one which relies on feeding back information to change practices, and the other which is a problem-solving approach to learning.

One approach to learning is to assume that researchers can contribute to the process of learning by feeding back information on how the logic models of the programme can be developed and by presenting research findings to practitioners and policy makers (see for example Judge and Bauld (2001). As Judge and Bald note, a necessary precondition for learning from feedback is the 'need to be *willing* to learn from feedback' (Judge and Bald 2001:36. Italics added). In our experience the usual setting for feeding back research findings is to a multi-agency partnership board or steering group and such willingness is not necessarily present. Listening to researchers present findings is essentially a passive activity which partners can choose not to engage in. It is also unclear that fora for stakeholders are necessarily conducive settings for learning and developing practices. It is most likely that the evaluation will have found changes, some of which may be positive whilst others will be negative, leading to feelings of disappointment and sometimes recrimination between partners (Weiss 1997a, 1997b; Brown 1995). Public settings and pressures to succeed inhibit any honest account of errors made, and inhibiting learning. All too often stakeholders start to 'blame' each other rather than taking responsibilities for shortcomings themselves or finding a resolution to the situation. The potential for researchers to promote learning and the development of new practices is, in these situations likely to be limited.

Another approach to learning, and one which aims to maximise the potential for learning during an evaluation, draws on an understanding of how learning takes place from Popper's thesis which is articulated by Burgess (2000) as

> …'all learning starts with problems, in particular those which arise when our knowledge bumps up against our ignorance. When we find that something is not as we supposed, or that there is something we do not know how to do, and determine to do something about it, we are ready to learn' (Burgess, 2000:54).

According to this theory when we learn something new, we come across new areas of ignorance, and the problem-solving process continues with the formulation of new problems and a search for solutions by trial and error. As Popper wrote:

> 'It is part of my thesis that *all* our knowledge grows *only* through the correcting of our mistakes. For example, what is called today 'negative feed back' is only an application of the general method of learning from our mistakes – the method of trial and error' (Popper, 1969:ix).

This model of learning informs our own approach to research. Researchers can facilitate learning by encouraging practitioners to be reflective and critical of their work and can enable practitioners to find their own solutions to difficulties they are experiencing in their everyday work. It is appropriate for researchers to challenge the assumptions informing practices, particularly where they are discriminatory and it is the active participation of practitioners in reflecting on their work brings about the potential for change because it provides the opportunity for problem solving and new learning. The task of the researcher is to use findings and examples from the research to encourage a reflective, critical problem solving approach and their role is to stimulate practitioners to problem solve, find their own solutions, and to develop new ideas. In this way learning is active and closely linked with changing practices since the new knowledge solves every day problems.

Thus, we would argue that there is more potential for learning where a researcher acts as a critical friend. That is, a researcher encourages reflexive and critical thinking, and encourages active learning by a problem solving approach. In this way evaluators are likely to be more effective at bringing about improvements as the learning is integrated into the everyday working practices and policy problems faced by policy makers and practitioners. This may involve challenging the status quo, questioning current practices of agencies, examining the value of the intervention itself and identifying errors which are apparent from the data collected and collated by researchers in their task of assessing the impact of an initiative.

Best practice

Policy-makers and practitioners can use information on best practice found in evaluations to encourage practitioners across the country to replicate the programme or project, and develop policies. This enhances the value of the research and increases the potential of the research to contribute to social improvement. For these reasons the notion of transferring practices and programmes should, in our view, be integrated into the evaluation framework from the outset. In this way researchers create hypotheses and carry out fieldwork with the intention of understanding good practices and how they may be transferred, rather than just inferring them from findings after the research has concluded.

The difficulty of transferring good practice is that of generalising from research which has taken place in one particular area, a problem which is relevant to all types of evaluations: 'As with other evaluations, TBE [theory-based evaluations] results cannot be mechanically applied to other sites' (Weiss 1997a: 513).

Popper identified the central problem of replication when he noted 'things may be similar but not the same' and it is not possible to strictly replicate successful projects based on 'mere similarity' - as things which are similar are only similar in certain respects (Popper 1968; 420-21). This point has been elaborated upon by Pawson and Tilley in their book on realistic evaluation, and they demonstrate that it is a complex task identifying which aspects of a project are best replicated; they note the general lack of success of replicating projects and how similar projects have been successful in one location but not in others (Pawson and Tilley 1997;127-35).

The complexities of replication are even clearer when we use 'situational logic' as a tool for analysing and interpreting data (Popper 1969). Situational logic assists us understand decision-making processes and therefore the processes of change; it recognises that according to the logic of their situation people and institutions pursue certain goals or aims and they do this by assessing which is the best way of achieving these goals within the given situation. Jarvie (1972) draws our attention to the possibility that a person or institution may find they need to choose between several different options in their pursuit of certain aims. As researchers we are interested in why particular decisions, rather than others, are made. Jarvie (1972) also highlights the significance of meanings and emotions in analysing the logic of the situation and understanding why particular decisions are made. As evaluators we are interested in how emotions, feelings and beliefs change as a result of the interventions of a programme. As Jarvie wrote:

> 'For social change, I believe, comes about primarily because people revise their beliefs about the world and society and on that account change the way they act'. (Jarvie 1972: ix)

Whilst using situational logic to analyse data illustrates how decision-making can be deeply embedded in local contexts and particular situations, it also enables researchers to find out about the beliefs of policy-makers, practitioners and service users and why certain decisions are made, rather than others, in response to a social programme.

If, however, evaluations are to contribute to social improvement in any general sense, then researchers need to find out about the processes which lead to social change and to identify what it is about an initiative that brings about change in one area that can be transferred to other locations. Chen and Rossi argue that developing a causative theory is necessary for generalisations to be made and it is the information about how changes come about that can be generalised (Chen and Rossi 1992:3). Similarly Pawson and Tilley (1997) have argued that to replicate research findings in other circumstances it is necessary to establish a plausible context – mechanism – outcome configuration to ensure a project will 'work' in another context. Pawson and Tilley argue that the generation of more explanatory information makes extracting general principles from the data possible, and it is these principles or generative causal mechanisms which are transferable as they provide the necessary conditions for decisions to be made (Pawson and Tilley 1997: pp 120 - 123).

We would argue that a potential way forward is by identifying the mechanisms of change as these are the actions which explain the outcomes of a social programme.

These mechanisms which bring about change can be understood as principles which can be generalised. We conceptualise mechanisms as defined by Weiss. Her definition is: 'The mechanism of change is not the program activities per se but the response that the activities generate' (Weiss, 1997b: 46). The attraction of this definition is that it contains an action component which is essential for explaining how changes occur. The 'action' component – the causal mechanism – can be a change in feelings, emotions or beliefs, in interpretations and understandings of situations, not simply a change in behaviour. Having identified change mechanisms researchers can then identify the practices which create the conditions necessary for the changes to occur. We would argue that it is these practices or practice principles which can be transferred, rather then projects in their entirety. The context is, of course, important in influencing the success of any intervention but at the heart of any replication are the practice principles; if an initiative has any chance of succeeding in a range of different locations then the causal mechanisms which are activated by effective practices, need to remain active whatever the context.

The idea of transferring practice principles rather than 'successful projects', can be illustrated with reference to some recent research we undertook on youth work schemes aiming to reduce offending and those at risk of offending in socially and economically disadvantaged communities. The Youth Inclusion Programme (YIP) was a replication of three successful projects (Brown, 1998). The aims, objectives, and outcomes of the initial three successful projects were replicated in 70 new areas and funded by the Youth Justice Board (YJB). The framework of working was also replicated; that is the targeting of 50 young people who were offenders or at risk of offending, locating projects in disadvantaged areas, the use of youth workers to deliver the programme for young people, and having a target of working with each young person for 10 hours a week, in order to include these marginalised young people into their community. The framework within which they operated was prescribed, and closely regulated and monitored by programme developers. Performance was also assessed by national evaluators, who contracted regional evaluators to ensure that the activities and outcome measures of each YIP were recorded on a computerised information system. This information was regularly up-dated to provide an ongoing record of performance.

The national evaluators, Morgan Harris Burrows (MHB), used this information system to assess the progress of the YIPs and the findings revealed that assessed against the outcomes measures of reducing crime, arrests, truancy from schools and school exclusions, the results were mixed. For example, arrests for the 'top 50' young people decreased in 61 projects between 1% and 117% and in six projects they increased between 1% and 24%. Neighbourhood crime decreased in 20 project areas and increased in 36 others. The decreases were up to 70% and increases up to 80% for the two-year period the YIPs had been operating[viii] (MHB, 2003).

In the example of the Youth Inclusion Programme the procedural and administrative framework was transferred rather than the principles or generative causal mechanisms. In our role as regional evaluators we incorporated into our interview schedules questions to find out about mechanisms of change. Through interviews with youth workers and young people we developed a model of good practice which were principles for youth workers working with young offenders (Ahmad et al, 2003). These principles were trust, respect, motivation, and self-sufficiency.[ix] They were the necessary conditions for decisions to be made by young people which could lead to the desistance from offending, they enabled young people to chose alternative actions to illegal activities. Whatever the context, they are the necessary practices to bring about the intended changes, namely to reduce crime.

Concluding reflections

The difficulties associated with undertaking a rigorous evaluation cannot be under-estimated and the purpose of this paper has been to raise four issues which, in our view, are integral to developing a robust evaluation approach. It has been argued that the inductive approach uses evidence as 'proof' which is less robust than a Popperian approach of using data to refute underlying assumptions and beliefs of a social programme. The Popperian perspective also alerts policy makers to the limits of an evaluation, that research findings are retrospective and situation specific. Researchers therefore need to be cautious about predicting that successful community initiatives will work in the future.

It is suggested that the more conventional approach of assessing the extent to which a programme has met its aims is broadened to include gathering information on the social problems themselves to find out if the policy is able to alleviate the problems it was designed to address. An example was given of where the causal problems of disadvantage within a local Sure Start area were not the same as those formulated by the national policy, with the implication that the main problems are likely to persist locally. This example therefore demonstrated how the incorporation of a problem-solving approach within an evaluation framework provides policy-makers with a better understanding of the longer term effect of the policy or programme.

In this paper we concur that evaluators can have a role in contributing to the implementation of community initiatives. It is argued however that evaluators can contribute more towards social improvement or social betterment by being a critical friend than by loosing their independence and becoming integral to the development of the programme. This position utilises a model of learning which is based on Popper's thesis that knowledge occurs through the correction of errors.

Finally it is argued that an effective evaluation framework is one which includes the identification of mechanisms of change or causal mechanisms so that it is possible to transfer good practice principles to other disadvantaged areas. To achieve this it is necessary that this approach is integrated into the initial evaluation framework, which again moves beyond the confines of assessing the outcomes of a programme in relation to its aims. It also encourages policy-makers to move away from 'rolling out' successful projects and to entertain the idea of replicating particular practices which may enhance the performance of a range of different types of social programmes.

References

Ahmad, E. and R. Knight and A. Parmar and A. Sampson and S.Themelis (2003) 'YIP Youth Work', Proceedings from a seminar, 21st June 2003. Centre for Institutional Studies, University of East London.

Brookes, N. (2002) *Monitoring performance: A survey of satisfaction with family support services amongst the Sure Start target population*. Barts and The London Queen Mary's School of Medicine and Dentistry. Unpublished manuscript.

Brown, P. (1995) 'The role of the evaluator in Comprehensive Community Initiatives' in Connell, J and A. Kubisch and L. Schorr and C. Weiss (eds) *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*, pp 201- 225. Washington: Aspen Institute.

Brown, S. (1998) *Youth Works: An Evaluation*. Faculty of Law, University of Sheffield, UK. Unpublished manuscript.

Burgess, T. (2000) 'The logic of learning and its implications for high education', *Higher Education Review* 32 (2):53 – 65.

Burgess, T. (2002) 'Towards a social science: a comment on Karl Popper's twenty seven theses', Paper presented at the Karl Popper Centenary Congress, University of Vienna, Austria: 3-7 July.

Chen, H-T. (1990) *Theory-Driven Evaluations*. London: Sage.

Chen, H-T. and P. Rossi (1992) 'Introduction: Integrating Theory into Evaluation Practice', in Chen, H-T. and P. Rossi, (eds.) *Using Theory to Improve Program and Policy Evaluations,* pp. 1-11. New York: Greenwood Press.

Connell. J. and A. Kubisch and A. Schorr and C. Weiss (1995) *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*. Washington: Aspen Institute.

Cordis Bright Ltd (2003) *Sure Start Abbey: Community Survey 2002/03*. Unpublished manuscript.

Fearnley, R. and N. McInroy (2000) 'Agents of change in East London: reflections on conducting urban regeneration research', in T. Butler (ed.) *Eastern Promise: Education and Social Renewal in London's Docklands*, pp 210-225.London: Lawrence and Wishart.

Foot Whyte, W.(ed) (1991) *Participatory Action Research*. London: Sage.

Griffith, J., and A. Sampson (2002) 'Using and developing the Popperian approach to evaluations', Paper presented at Karl Popper Centenary Congress, University of Vienna, Austria. 3-7 July.

Israel, B., Schurman, S., and Hugentobler, M (1992) 'Conducting Action Research: Relationships Between Organisation Members and Researchers', *Journal of Applied Behavioural Science* 28 (1): 74 –101.

Jarvie, I. (1972) *Concepts and Society*. London: Routledge and Kegan Paul.

Judge, K. and L. Bauld, (2001) 'Strong theory, flexible methods: evaluating complex community-based initiatives', *Critical Public Health* 11 (1): 19 – 38.

Kubisch, A. and K. Fulbright-Anderson and J. Connell, (1999) 'Evaluating Community Initiatives: A Progress Report'. Available at : http://www. aspenroundtable.org/vol2/kubisch.htm (site visited:15 June 2002).

Lincoln, L. (2001) 'Engaging Sympathies: Relationships between Action Research and Social Constructivism', in P. Reason and H. Bradbury (eds.) *Handbook of Action Research*. London: Sage.

Majone, G. (1980) 'Policies as Theories', *Omega: The International Journal of Management Science* 8 (2): 151-162.

Mackenzie, M., and A. Blamey, (2005) 'The Practice and the Theory: Lessons from the Application of a Theories of Change Approach'. Evaluation 11(2):151-168.

McQuail, L and V. Wigfall (2001) 'Euston Sure Start: A needs survey'. Available at: http://www.ness.bbk.ac.uk (site visited: 20 February 2003).

Morgan Harris and Burrows (2003) 'Evaluation of the Youth Inclusion Programme. http://www.youth-justice-board.gov.uk. (site visited: 30 November 2003).

National Statistics Online (2003) 'Neighbourhood Statistics'. Available at: http:// www.neighbourhood.statistics.gov.uk. (site visited: 3 February 2003).

Nutley, S., and I. Walter, and H. Davies (2003) 'From Knowing to Doing: A Framework for Understanding the Evidence-Into-Practice Agenda', *Evaluation*. 9(2): 125-148.

Oakley, A. (2000) *Experiments in Knowing*. Cambridge: Polity Press.

Parmar, A., and A. Sampson, and P. Jordan (2003) Final Site Report of Croydon Domestic Violence Advocacy Project. Submitted to the Home Office. Unpublished manuscript.

Pawson, R. and N. Tilley (1997) *Realistic Evaluation*. London: Sage.

Popper, K. (1968) *The Logic of Scientific Discovery*. Second edition. London: Hutchinson.

Popper, K. (1969) *Conjectures and Refutations: The Growth of Scientific Knowledge.* Third edition. London: Routledge and Kegan Paul.

Sayer, A. (1992) *Method in Social Science: A Realist Approach*. 2nd Edition. London: Routledge.

Reason, P. and Bradbury, H eds (2001) *Handbook of Action Research*. London: Sage.

Weiss, C. (1972) *Evaluation Research*. New Jersey: Prentice-Hall.

Weiss, C. (1995) 'Nothing as practical as good theory: exploring theory-based evaluation for Comprehensive Community Initiatives for children and families', in Connell, J. and A. Kubisch and L. Schorr and C. Weiss, (eds) *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*, pp. 65-92. Washington: Aspen Institute.

Weiss, C. (1997a) 'How can theory-based evaluation make greater headway?', *Evaluation Review* 21(4):501 -524

Weiss, C. (1997b) 'Theory-Based Evaluation: Past, Present and Future', *New Directions for Evaluation*. 76: 41 – 55.

---

[i] The Centre for Institutional Studies, Social Sciences, University of East London draws on the philosophy of Karl Popper to inform its teaching and research. My thanks are due to Emma Ahmad, Tyrrell Burgess, Jon Griffith, Mike Locke, John Pratt and Paul Robson for their comments on an earlier draft of the paper. In particular I would like to thank Tyrrell Burgess for commenting on a later draft and for discussing the referees' comments with me. The usual disclaimer applies.

[ii] Thanks to Tyrrell Burgess, Centre for Institutional Studies, University of East London for providing this banal example.

[iii] This apt expression was first used by Jon Griffith during a seminar at the Centre for Institutional Studies, University of East London.

[iv] Interestingly although the HAZ evaluation adopts a ToC approach, including an inductive analysis of the data (Judge and Bald, 2001) alternative explanations were explored. A referee commented that the hypothesis that installing fire alarms in homes helps reduce accidents was refuted by empirical evidence, refuting the chosen theory. This suggests that in practice an inductive approach is inadequate.

[v] We would claim that our findings are tentative and that our research aims to find the truth, rather than be 'right'. There were various 'messy' factors to take into account; for example some of the advocates were not working to the 'model' of advocacy espoused in the bid documents, and these considerations necessarily make for uncertainty.

[vi] The most disadvantaged domains, and overall ranking, are assigned 1 and then each domain within a ward and the overall score for a ward is ranked in ascending order with the least disadvantaged domains, and overall ranking being 8414.

[vii] Researchers may take on roles associated with action research (Reason and Bradbury, 2001) and participatory action research (Foot Whyte, 1991) which typically include problem identification and problem-solving with participants of a social programme. Whilst evaluators may take on some of these roles, the primary purpose of an evaluation is to assess outcomes and gaining access to information to perform this task remains a priority and some action research roles may jeopardise this purpose.

[viii] This figures do not take into account all 70 projects as several projects did not complete their databases (MHB 2003).

[ix] At the time of the research these general principles best explained the data. Of course 'good practice' models are themselves hypotheses which are tentative and can be falsified. Our research contract ended shortly after we developed the good practice model and therefore did not continue to refute them.