The case for an experimental approach in applied social research: an illustration from the area of civil renewal policy[1]

Gerry Stoker and Stephen Greasley
IPEG (www.ipeg.org.uk), University of Manchester, February 2005
gerry.stoker@manchester.ac.uk

The paper focuses on the case for an experimental research approach to be applied to the development and evaluation of policy in the area of civil renewal but the arguments of the paper could apply to other areas of research where the nature of policy and implementation demands a similar exploratory approach. Experiments involve the testing of propositions about "what works" through a process of trial and evaluation which runs alongside the development and implementation of a policy. In this paper we argue that design experiments can help in the shaping of a policy and that randomised controlled trails can support a judgement about their ultimate impact. In particular we claim we can use experiments to produce warrantable knowledge that in turn can match a set of scientific principles on which many researchers would agree (National Research Council,2002) As social scientists we would all value the production of evidence, the linking of evidence to theory, scepticism about claims to knowledge, a keenness to rule out alternative explanations and a commitment to constructive criticism as key ingredients in developing knowledge. The principles to guide scientific investigation are laid out in more detail in the appendix to this paper. Their aim and our aim is to provide guidance to encourage the most rigorous and systematic approach to the collection of evidence in respect of the development of public policy.

Faced with the challenge of developing systematic, theory driven and warrantable claims to knowledge the social scientist has two broad strategies available. One approach that

[1] This paper is a revised version of a paper produced in 2004 with financial support from the Home Office and its development draws on resources provided by the Home Office's civil renewal 2005 research programme which is run out of IPEG at the University of Manchester

has attracted a lot of support is captured by the term realistic evaluation as promoted by Pawson and Tilley(1997). It involves judging policies after their launch against a set of worked up theories of change embodied in the policy and revealed by the investigative activities of the researcher. Developing and testing theories of change in this manner means recognizing the embeddedness of all interventions in a wider social context. The mechanism for achieving change is premised on the bringing together of resources and a process of reasoning; doing this activity is summarised will produce this positive outcome. But realistic evaluation demonstrates how any mechanism is context dependent; it only works in particular circumstances. Understanding the mechanism and the theories of change embodied with it and judging the context and circumstances in which it is likely to work are central to the strategy of realistic evaluation.

The approach of realistic evaluation has much to commend it but it does tend to assume that the policy makers and practitioners will develop their policy responses or interventions in a muddled or unsystematic manner and it is up researchers to develop coping strategies. Such an assumption may be 'realistic' but it is not necessarily desirable. Moreover, despite all the hard work for researchers suggested by Pawson and Tilley to develop more explicit theories of change – drawn out from the policy makers and practitioners after the event - and the associated testing of those theories by an array of sophisticated research techniques, there is a sense that what is being assessed is post hoc rationalisation by policy makers and practitioners or, worse, the researcher's impressions of point of the policy that may have limited value. The theories of change central to the realistic approach are constructs, formed after the event, of what might work. They are often the products of the reasoning of researchers as much as those of policy makers or practitioners. The testing of those theories of change is of limited value because it is not explicitly tied into policy and practitioner processes and subsequent policy development. It is not the argument of this paper that the approach advocated as realistic evaluation can yield no valuable insights, but it to argue that there might be a better way if research, policy development and practitioner insight were better integrated.

That better way involves a second strategic approach to applied social science: a strategy of experimentation. Experiments involve the identification of causal mechanisms of change and their testing in particular contexts but crucially these activities run alongside the development and implementation of a policy. Two forms of experiments are promoted here. The first involves a form of action research in which researchers, alongside policy makers and practitioners, diagnose a problem, plan and organise action steps to respond to that problem and then seek to evaluate that the effectiveness of the intervention. The design experiments here, however, are not valued because they empower practitioners or because they enable practitioners and researchers to empathise but rather because they provide a stepping stone to plausible and testable propositions about what works. The second form of experiment advocated here-randomised controlled trials- can build on design experiments to produce a warrantable test of an intervention and judge whether it works and why.

The civil renewal agenda (Blunkett, 2003a and b) is a particularly appropriate one for the development of an experimental approach. For policy makers civil renewal presents a double puzzle. It is not clear what should be done. Moreover, civil renewal is obviously a policy area where government cannot do everything on its own. Can we discover the mechanisms that will get people engaged in defining and solving the problems of their communities in a sustainable way? Why does it appear that, for example, some schemes for citizen engagement seem to work better in some areas than in others?

Civil renewal is a policy area where we should follow the advice of Donald Campbell (1979), and see 'reforms as experiments'. Campbell goes further to suggest that such an approach is especially appropriate where policy makers and practitioners have justified reform 'on the basis of the importance of the problem, not the certainty of their answer, and are committed to going on to other potential solutions if the first trial fails' (Campbell, 1979, 109). It may be over-rationalistic and utopian to suggest that policy can be moved forward by sequential trials, but the spirit of Campbell's commitment to research running alongside reform would seem highly desirable in the area of civil renewal.

We need to move beyond a commitment to best practice guides in so far their findings rests on a series of case studies of ‘success’ cannot establish what policy works. First the case studies struggle to provide an explanation of why a particular initiative worked because there is no systematic framing of the causal factors that might be involved. Second the case studies cannot be replicated or matched against areas where different or no intervention took place. Case studies are by definition stories or narratives. They are written from experience, provide interpretative frameworks and contribute greatly to our understanding. But the empirical method in science has different features and makes different demands. It insists on codified procedures for recording observations, mechanisms for weeding out bias in data collection, ways of testing evidence against propositions and a commitment to replication and generalization

The main thrust of this document is to suggest that it will be necessary to develop a framework for evaluation that moves from a design, formative phase towards a definitive and long-term judgement about the impact of an intervention. It makes the case for design experiments to explore policy and implementation options and randomised controlled trials to establish impacts in a warrantable manner. The paper outlines both options before illustrating their application in the field of community housing.

Design experiments: A method of exploration

1. The design experiment approach

‘Design experimentation’ or ‘design based research’ takes inspiration from an ‘engineering’ understanding of social and political interventions rather than a ‘natural science’ viewpoint. As yet the approach has focused primarily on educational issues, particularly the design and use of teaching strategies, instruments and curriculum development but it would appear to have a wider application.

In the initial exploratory phase in developing and crafting initiatives we propose to use the methodology of design experiments. It is worth just emphasising the understanding of the exploratory phase implied here. The aim is to actively search for what might work, to look at options, potential improvements and interventions. This is a phase that requires an iterative and sustained relationship between practitioners, policy makers and researchers. It is assumed that in this phase and in the latter phase focused on impact that the researcher is interested in both whether and how an intervention is working. A commitment to understanding a theory of change runs throughout both phases (Chen, 1990; Pawson and Tilley, 1997) because it is essential to the achievement of the scientific understanding. We need approaches that can help explore impact but also help us understand why change has occurred.

Plainly in a precursor to a design experiment the researcher., along with practitioners and policy makers would review the available evidence of what works in as systematic manner as possible. A systematic review is governed by three factors that distinguish it from a straight-forward literature review. First the search is more rule-driven and rigorous. There should be an explicit statement of collection rules and an attempt if possible to cover all published material and other evidence in non-published forms if it can be found. Second there should be transparent criteria for appraising the studies and the quality of evidence on which they are based. There should be some methodological quality threshold before a study is included in the review. There are a number of questions to be asked in critically appraising the literature. A helpful checklist for guiding systematic reviews is provided in Magenta Book (2003, Ch 2, Appendix 1).The third general issue to be addressed in systematic reviews is that there should be an explicit attempt to establish criteria against which studies can be compared and how a judgement is to be made about the cumulative impact of the research. The choice here is between a narrative and a vote counting account. The latter option may be less available in the field of civil renewal because the available evidence is unlikely to allow such judgements to be easily made. A narrative option is the most likely one to be feasible but it is essential that

the criteria for making judgements and comparisons between the different findings is based as clear and explicit statements. Without such clarification it is not possible to judge whether the systematic review has been subject to hidden or latent biases.

Neither systematic reviews nor modelling will necessarily deliver comprehensive 'on the ground' insights that may be crucial to developing civil renewal policy in this exploratory phase. The nature of the interventions is likely to be so complex and uncertain and the detail of the causal factors at work so difficult to fathom that there will be a case for developing pilot schemes as a core part of policy development. This situation is where design experiments come into there own as a form of piloting.

There are three main features of a design experiment. First a cycle of interventions is applied to a real world context and detailed records are kept of the process of 'enactment'. At the end of a cycle the data is analysed, modifications are made to the design of the intervention and the process is repeated. The approach is therefore experimental in an everyday sense of the word rather than in the manner of a controlled trial. Researchers intervene in real world settings to assess the impact of the intervention, adjust the design of the intervention and continue repeatedly until some judgment is made that the intervention is working satisfactorily, or is unlikely to work satisfactorily.

Second , because the design of interventions is guided by theory, design experimentation is more than a trial and error approach to discovery. It aims to produce models as well as a successful interventions. The design of an intervention is based on clear theoretical presuppositions which are placed under scrutiny as the intervention proceeds. The idea that theory development is essential to evaluation is now widely accepted in the evaluation literature (Chen, 1990; Pawson and Tilley; Magenta Book, 2003, Ch 1). The overarching aim is to identify the logical sequence through which an intervention or mechanism might produce effects. Such 'a theory of change' could be established through interviews with key policy-makers and practitioners and through participant

observation by the researchers. In addition as part of a design experiment it might be possible to carry out a simulation exercise with key participants in order to work through a process of change in an artificial setting but one informed by an understanding of the context in which the policy is being developed. Focus groups might give a similar mechanism for testing out ideas. Finally, there are paper-based exercises to map out connections or even the possibility of computer simulations.

Thirdly, theories are evaluated not against some standard of truth value (whether absolute or comparative), but rather on their contribution to the successful design of interventions. Designs (and policies) are evaluated more instrumentally against some measure of utility. The specific design may not apply in another setting because of the specifics of context and circumstances, but the general thinking and underlying theory might. Design experiments are about refining interventions so that they work and also about trying to establish some more general understanding of what underlies the achievement of an effective intervention. The idea of design experimentation is not about a final evaluation but rather a process of exploration that links researchers, practitioners and policy makers. Its claim is to allow for the adaptation and development of interventions and to fine tune them to meet the objectives and challenges of policy and practice.

Depending on the programme being researched the cycle can last weeks or years. A common characteristic of design experiments is their collaborative nature and the active involvement of practitioners in the design and implementation of research. As such the approach recognizes that practitioners always have valuable information and insights that can help the successful development of policy as a whole and that policies rely on this knowledge for their successful application in the variety of different contexts in which they are applied.

A key question for design experimenters is how to decide when an adjustment to an intervention goes beyond a critical boundary so that the intervention no longer accords with the underlying theory. There is a tension here for design experimentation. Scientific standards highlight issues of validity and reliability in the assessment of knowledge, from

the engineering point of the view the critical question concerns catastrophic 'failure' and avoiding it. Design experiments are about fixing things and adapting. They lend themselves to exploratory phases in policy development but are not appropriate on their own for other phases.

2. Design experiments: strengths and weaknesses

Pitfalls can arise from the collaborative nature of the research, either because of the actions taken in implementing the intervention or because of the way in which data is recorded and analysed. In education research, teachers (and students) who know they are involved in a piece of research may, through enthusiasm for the project, work harder than otherwise (the Hawthorn effect). The teacher or school may select an unrepresentative group of students to participate, altering the chances of an intervention appearing to be successful. Furthermore, because design experiments are a type of action research, researchers find themselves in the roles of both advocate and judge. This can lead to a bias towards positive assessment of an intervention (the Rosenthal effect). The usual procedure for mitigating measurement bias is to ensure that the researcher does not know which group has received the treatment. This is not an option for design experiments because of the intensive and often qualitative nature of the data collection.

Design experiments are conducted in complex contexts involving many different variables all interacting with one another. In the absence of the systematic screening provided by randomised control, any attempt at causal inference faces serious problems in disentangling all these factors. Successful interventions may be a result of the additional resources or time put into an activity rather than anything to do with the nature of the intervention itself, or it may be a result of the unrepresentative nature of the group that participated in the intervention. In short design experiments can provide a useful element in discovering applicable and warrantable knowledge but they will not do the job on their own. Design experiments, as both McCandliss et al (2003) and Shavelson et al (2003) argue, can benefit from appropriate collaborations with other research approaches.

It is clear, however, that the design experiment does have something unique to offer. Firstly, in the piloting stage of a policy, the design experiment cycle (theoretically informed design, enactment, systematic assessment, redesign, enactment, assessment), presents an opportunity to illuminate a number of the key trade-offs in policy design. In the context of civil renewal design, experimentation can help distinguish core policy aspects, where policy needs to be prescriptive and uniform in implementation, from the local choice aspects, where variation in local context implies reliance on the local knowledge and experience of the professionals involved in developing programmes on the ground. Design experiments at this early stage may also aid the development of evaluative criteria for a policy that do not lead to perverse incentives and administrative overload.

Secondly, design experiments satisfy the impatience of policy makers for something to be done, but they do so in a way that informs and develops effective policy interventions rather than relying on a leap of faith. Design experiments do imply intense data collection and long periods of research activity and of course there are costs associated with these activities. This may be the cost of better policy: by carrying out research and policy adaptation simultaneously, design experiments can provide nearly immediate payback from research activity. The unique benefit of this approach is that it integrates practice with systematic knowledge development.

Thirdly, in the dissemination and implementation of policy, design experiments can help professionals to adapt programmes to local needs and can help with a process of continual refinement, improvement and change in the face of changing needs. By linking policy dissemination to the underlying theory which informs the policy, design experimentation has the potential to be a more powerful instrument for improving delivery across the country. It is also likely that design experiments as dissemination can facilitate local ownership of programmes because they assume a more autonomous and proactive professional, than the passive consumer of good practice guidance.

A final benefit would be an increase in practice informed by research and in practically relevant research. This is not to argue for turning practitioners into researchers; there is still a role for experience and professional judgement in policy implementation. Nor is it to diminish the role of basic research in academia. Rather, it is to argue for a more systematic route between research and real world practice.

Randomised controlled trials: establishing impacts

1. Randomised controlled trials: judging impact

The exploratory phase can reveal many powerful insights and enable interventions that work to be better understood. Ultimately the goal of evaluation should be to show that some interventions have made a positive difference greater than could be achieved by alternatives. To answer the 'what works' question means being able to demonstrate the impact of an intervention, what effect was achieved and how that effect was achieved. Again, a variety of evaluation methods are available but particular attention is paid here to the prospects of understanding cause and effect and establishing net impact effects using definitive randomised controlled trials (RCTs).

At the stage of long-term evaluation the intervention will need to be held as a steady and particular mechanism if it is assessed in a way that avoids complications. The defining feature of a controlled experiment is the establishment of two or more comparison groups that are similar in all respects and which may affect the outcome of the process being studied (Gomm, 2004, chapter 2). Importantly, in true experiments it is the *random* allocation of research subjects (be they people, groups or institutions) to one or other group that is used to ensure that the groups are similar in ways relevant to the research questions. These groups are then treated the same in all ways except those that are the focus of the research. Finally, observations are made on key variables for each member of the two groups both prior to and after the intervention. 'Pre' and 'post' intervention measurements are the minimum required. In practice many research projects with an

experimental design also monitor the implementation of the policy action using a variety of qualitative and quantitative methods.

The strength of random allocation to different groups is that it can control for the influence of factors that are known to affect the outcome *and* the influence of factors that may affect the outcome of the trial but are *unknown* to researchers. Given the complexity of the civil renewal agenda as a research area it is likely that in many instances there will be unknown factors, or as importantly factors that cannot be measured, which will affect the outcome of interest. But the major caveat here is that randomisation will only work to produce similar groups if the number of subjects being randomised is large enough. This may be a particular problem for civil renewal where the unit of analysis is often at an aggregate level, such as a community or neighbourhood. In short, where there are many possible factors (other than the intervention or policy) that can have an affect on outcomes of interest, then randomisation needs a large sample, but where outcomes are thought to act at community level there may be fewer subjects available to randomise.

The Medical Research Council has developed guidance notes for cluster trials where randomisation occurs at a level of aggregation higher than the individual. Cluster randomised trials (CRTs) are used to evaluate group interventions and individual interventions where there are group level effects (MRC, 2002). In CRTs social groups rather than individuals are randomly allocated to intervention and control, although in medical research the outcomes of the intervention are still normally measured at individual level. This strategy is appropriate in a number of circumstances. It can be used where interventions are delivered to a group and affect the group as a whole; where an intervention is aimed at a professional or an administrative unit and the interest is in the impact on the community served; and where there are spillover effects - for example when an intervention given to one individual affects others in a group (MRC 2002: 3-4). This last condition is particularly relevant for the civil renewal agenda, where spillover through social networks is likely to be an important factor. Some technical problems with this type of design have been identified (MRC, 2002) but they can be addressed with the

correct statistical procedures. Another concern is that the two stage recruitment process may lead to bias in the selection of the cohorts.

If there are too few subjects available for study, complete randomisation cannot lead to powerful results and some form of stratification is likely to be used to try to ensure that intervention and control groups are similar in known characteristics. The research then moves into quasi experimental strategies such *matched pair* designs. When there are few subjects, the unavoidable trade off is the deliberate controlling of 'knowns' and the random controlling for 'unknowns'.

The strength of controlled experiments is in identifying causal relation between intervention and outcome at work in the cases studied. They can produce results with strong 'internal validity', but policy makers and practitioners also need to know how widely the results and the policy can be generalised. If the experiment is large enough, sub-group comparisons can be drawn on to infer whether an intervention works better depending on gender, location, class or a variety of characteristics. But there is no easy answer to 'where and when it works' other than to continually refine research questions and carry out more research, some of it using controlled experimentation and some of it using other methods. The key goal for researchers, policy makers and practitioners alike must be to ensure that the results of this research are cumulative.

Without prejudging policy preferences, it is possible to draw on some of the experimental work in political science that exists to show what kinds of research might be possible (for a review see McGraw, 1996). Donald Green and Alan Gerber at Yale University (see www.yale.edu/isps/publications) have launched a number of experiments mostly directed towards testing how to persuade people to get out and vote. They have shown in a number of studies using random assignment of an intervention, that it is personal contact that has the greatest impact in encouraging people to be active (Green et al, 2003; Gerber and Green, 2000, Green and Gerber, 2001). They have also suggested that experiments could be constructed to examine claims made by Fishkin (1995) for deliberative polling with the use of discussion forums prior to the taking of votes or Putnam's (2000) claims

about the impact of social capital on political participation (Green and Gerber, 2002). Studies using experimental techniques have also been used to test the impact of various types of citizenship promotion messages. The impact of a similar initiative to encourage civic engagement in several locations, but with subtle variations in place, would be possible (see American Behavioral Scientist, autumn 2004, edited by Green and Gerber for a critical discussion of the possible applications of randomised experiments).

2. Strengths and weaknesses of RCTs

The key strength of RCTs is that if the conditions of internal validity are sustained, then the question of the impact of an intervention can receive a definitive answer. By random allocation of units to be effected affected by the treatment, and by comparing the impact of groups that have been treated with control groups that have not, and if all other things have been kept equal between the treated and non-treated groups, we can see whether the intervention has worked. This is the core logic of experiment. The difficulty lies in maintaining the conditions for effective implementation and internal validity. As Jowell (2003, 16) notes, RCTs have been widely used in Britain in medical research and have recently been carried out in a few cases in applied social interventions, mostly in the employment and training field. He goes on to note: 'It is fair to report that most of these pilots were bedevilled by practical problems of implementation' (Jowell, 2003, 17). The problems were in part explained by the inadequate training and support given to the staff involved, and to the relative novelty of the approach, it is claimed. Conducted well and effectively, RCTs are capable of producing the highest quality of warrantable knowledge.

Practical difficulties affect RCTs, as they do all other research methods, even in circumstances where there is considerable experience of implementation. Greenberg et al (2003) review the long history of RCTs in the United States in the social field and provide examples of where trials have had to be aborted or modified because of administrative and other problems. Particular problems can be caused, for example by attrition between participants from different programmes, although the study shows that problems can also be overcome in most instances.

Another review of U.S experience (Moffit, 2003) in welfare programmes argues that although challenging issues relating to the internal validity of experiments can generally be overcome, there are greater challenges in establishing external validity. Controls over the treatment and the groups in the trial can be established to a sufficient standard to show that the effects found in the particular trial are valid. The bigger set of difficulties revolves around establishing the external validity the findings. Will the intervention work in the same way elsewhere? There are a number of commonly raised problems which would have to be considered in the context of work on civil renewal. Estimating the effects of system-wide reform can be particularly problematic using randomised trials because of contamination effects. Feedback mechanisms (such as networking between people and media coverage), perhaps caused by the intervention, may in turn make it difficult to maintain the purity of the trial. If a scheme is being rolled out nationally it is very hard to ensure that the control groups are not in some way affected by the intervention. If the scheme is kept very local and specific there may be particular site or place effects that are difficult to control for, and there is always the danger of unplanned and uncontrolled treatment variation. If the intervention is complex in its potential causal factors, then knowing what is going on inside the 'black box' can be difficult. The design of an experiment with enough variation built-in it to allow for the testing of a range of sub-factors in an overall framework of causality is a considerable technical challenge in terms of policy design and with respect to maintaining its legitimacy and acceptance. None of these difficulties are insurmountable but their existence needs to be taken into account in the design of experiments and in deciding the balance between experimental and non-experimental forms of evaluation.

Pawson and Tilley (1997, Ch 2) present a critique of the underlying model of social causation that, they argue, underwrites the experimental method. They argue that experiments rely on a 'successionist' theory of causation: that causation is established by following the connection between cause and effect. The aim of the experiment is to clear the extraneous noise out of the way so that a clear connection can be drawn between an intervention (cause) and an outcome (effect). Pawson and Tilley draw on a generative

theory of causation to argue that the real challenge is to view social interventions internally and understand how and why social programmes have the potential to cause change. Change is a product of human action and the role of the evaluator is to understand the conditions under which change can occur. We would counter experiments that could contribute to an understanding of causation as favoured by Pawson and Tilley and are not dependent on a 'successionist' logic (Bennett, 1996). In short understanding causes could be integrated into the experimental method without great difficulty.

A range of other doubts have been raised about the ethical and political viability of RCTs. Jowell (2003, 17) notes significant political concern in the UK over the random allocation of individuals to treatment or non-treatment in respect to some of the employment and training pilots. In short, is it fair that some people should receive help and others not? The reply is, of course, that it is only fair if we are not sure whether the intervention will work. Then it can be argued on utilitarian grounds, to do with the greater good, that experimentation to find out what works is ultimately of benefit to all. Notwithstanding this riposte, the legitimacy of any RCTs will have to be sustained by careful communication and explanation and by showing how they are part of a wider programme.

There are also wider issues about whether RCTs are really suited to the UK parliamentary system and process of policy development (Hogwood, 2000). The relatively widespread use of social experiments in the United States, it is argued, reflects the 'natural' focus of states as sites for experimentation. The long-standing tenure of many senior legislators at both federal and state level means they can wait for results, and the generally decentralised focus of policy making in turn means that policy does not have to be right because the real issue is the competition between different mandates and options. The UK in contrast has politicians who tend, because of their ministerial careers, to focus on the short-term. They operate in a system which is relatively centralised (notwithstanding devolution) and adversarial in style, so that policy options, once adopted, tend to be defended and cannot so easily be subject to trial. Moffit (2003), however, suggests that

the decentralised structures of the United States make experimentation more difficult as individual states or local jurisdictions go their own way on trials and do not follow national guidelines or protocols. It may be that the more centralised systems of the UK would suit experimentation if the issue of short-term pressures to deliver can be kept in check.

The final challenge that RCTs have faced is whether they will deliver enough in order to justify the scale of investment that might be required. Although some are sceptical about how much the tradition of experimentation has delivered in terms of definitive results (Pawson and Tilley, 1997), others are more positive. A detailed review of the experience in the United States of social experiments (Greenberg et al, 2003) shows that the impact of the research in part depends on its quality but also on its timeliness, effective communication and wide applicability and relevance to the concerns of policymaking. This indicates that where the technical, practical and other difficulties can be overcome, RCTs can make an effective contribution to policy. The judgement of Jowell (2003, 19) is that while RCTs are not the be-all and end-all in methodological terms they are 'seriously underused in Britain in circumstances where the technical advantages would seem to outweigh their other potential difficulties'

An experimental research strategy: an illustration

To illustrate the potential of an experimental strategy the study of community ownership schemes in Scotland will be used (Clapham and Kintrea, 1994) The first schemes were set up in 1986 and 1987 in response to pressure from residents for improvements that the local council could not provide. By 1995 almost 13,000 houses were operated by 40 Community Ownership Organisations in Scotland. The schemes were promoted under the general title of community-based housing organisations (CBHOs) and rested on the principle that all residents have the right to be involved in the management structure of their community. In particular under the schemes residents were given full responsibility for the housing stock which they collectively came to own in some form of mutual co-operative or community-based housing association. They were responsible for its

development, improvement and management. CBHOs operated on a relatively small scale with each organisation taking control of between 100-400 properties, primarily within the context of larger estates.

The detailed appraisal of the schemes by Clapham and Kintrea is interesting because it contains data collected in the late 1980s and data collected in the mid-1990s. The research suggests that CBHOs have had important positive impacts. The outcomes reported in the research are encouraging but not conclusive. Clapham and Kintrea in their early study report some positive results comparable to those achieved by CBHOs for other forms of housing management including a tenants management co-operative and a locally-based community housing association. From the work that has been done, it is not so clear which particular contextual factors support successful outcomes, which specific levers make a difference and which of the various processes of community organisation and activity are central to the development of sustainable schemes.

An experimental approach could have addressed the limitations of the evidence produced by Clapham and Kintrea. A design experiment could have piloted the idea of the CBHO institutional form in twenty locations. The locations would contain some variation (e.g. urban, rural, suburban) to allow consideration of the impact that context might make on the scheme. Design features such as organisational structures, governance processes and degrees of support could be varied and tested in an iterative process between practitioners and locally-based researchers with lessons and insights being shared and more systematically investigated by an over-arching research and policy team. Better measurements of the outcomes that are desired could be developed or refined. The time necessary for the design experiment phase would vary and to some extent should be subject to some pragmatic judgement, as the idea is to take forward our understanding of the policy and its research assessment, while not assuming that some final point of perfection will be achieved. The goal is to achieve both a policy and an ultimate research design more advanced than before.

Experiments and interventions in order to be judged need to be theoretically framed. In other work the researchers (Lowndes et al, 2004) have produced an over-arching model of factors underlying participation. Table 1, below, sets out the CLEAR theoretical framework to explain levels of participation and then uses that framework to explore and test various design options.

According to the **C.L.E.A.R** framework people participate when they **can,** when they have the resources necessary to make their argument. People participate when they are part of something; they **like** to participate because it is central to their sense of identity. They participate when they are **enabled** by an infrastructure of civic organisations. People participate when they are directly **asked** for their opinion. Finally people participate when they experience the system they are seeking to influence as **responsive.**

The introduction of this theoretical element means that there can be significant knowledge transfer between community housing and other elements of the civil renewal policy agenda. Understanding the underlying causes of the effects observed in one case can provide the basis for better understanding and effective design of other interventions in other settings. This in turn has considerable potential practical benefit.

Table 1: Promoting participation: from theory to design

Participation factors	**Associated policy target**	**Design experiment**
Can do: skill, education and other resources to support participation are present	Capacity building	**Research questions:** what is the best way to provide the skills needed by citizens? Would these techniques vary between social groups? **Possible areas of investigation:** • Learning by doing • Formal courses • Literature • Interactive computer based learning • Peer-group led learning
Like to: sense of engagement in the community is experienced	Sense of community, civic engagement, social capital and citizen-ship	**Research questions:** how can public action enhance sense of community? **Possible areas of investigation:** • Interaction between CBHOs and schools as part of civic education • Emphasising the skills/experience that can be gained • Design of websites, posters or newsletters • Community events
Enabled to: network of umbrella groups exist to support participation	To build the civic infra-structure so that there are groups and organisations around to channel and facilitate participation	**Research questions:** can membership of social/leisure groups lead to more civic engagement? **Possible areas of investigation:** • Does current public support for groups enhance or detract from community cohesion? How can public support be designed to enhance cohesion? • Allowing some of the funds for community activities to flow through CBHOs
Asked to; people are directly asked to participate	Mobilisation schemes that are diverse and reflexive	**Research questions:** what is the best way to disseminate information about opportunities for activity? **Possible areas of investigation:** • Encouraging recruitment through social networks – bring a friend • Designing institutional links between social or sports activities and CBHOs • Designing websites, newsletters, community radio
Responded to: when they do participate people see they are listened to	A public policy system and institutional environment that can show a capacity to respond	**Research question:** what features of institutional design affect the likelihood of citizens becoming involved? **Possible areas of investigation:** • publicising responsive action • methods for ensuring fairness • keeping citizens informed • introducing more flexibility into accountability and funding structures • reducing the number of steps to a decision

After the initial insights from design experiments our understanding of the scientific context is much clearer and our measurement instruments are more sharply defined. The policy is ready to be launched more widely but hopefully in a way that allows a commitment to investigate its effectiveness in a rigorous manner.

In a randomised controlled trials (RCTs) the refined CBHO intervention could be contrasted with sites of non-intervention. The research that Clapham and Kintrea carried out on CBHOs in Scotland convincingly demonstrates that there was a positive effect on some important outcomes. It is difficult however to identify exactly the reason for this improvement and the likelihood that similar structures would lead to improvements in other conditions. For example, the CBHOs were set up in response to the demands of the community. This implies that there was already some organising capacity and community feeling. Would CBHOs work in areas without these conditions? Would the organised communities that pushed for CBHOs have found other ways of improving the management of their housing if the reforms had not been forthcoming? A randomised control trial can help to disentangle some of these factors.

The first stage in developing a RCT is to specify the question. This will depend primarily on the concerns of policy makers. Some questions might be:

- Does a CBHO lead to improvements in satisfaction with housing management?
- Does a CBHO have positive effects on civic engagement in an area?
- Does a CBHO have positive effects on how residents perceive their area?

Before the questions can be addressed the relevant comparator has to be identified. The effect of CBHOs can be compared with other local authority provision or with other methods for tenant involvement, depending on the state of knowledge from previous research. CBHOs are reforms delivered at community level, hence the research will take the form of cluster randomised trials. There may be some stratification required to control

for known variables such as ethnic mix. One possibility would be a series of matched pair comparisons, to try to measure the effect of CBHOs in different contexts. Once similar pairs have been identified one will be randomly assigned to have a CBHO established and the other would act as comparator. Measures would be taken on a number of variables of interest ideally by researchers who do not know which area has the CBHO. Researchers would collect a record of the implementation of the CBHO, its outputs and any other programmes which might impact on the outcome variables. A similar record will be required in the comparator cases. Measurements are then taken after the period and differences between the control and the intervention groups are used to assess the impact of the reform.

A number of key decisions need to be made by the designers of the evaluation – for example how many cases to include, how long a period the comparison should go on for – these have implications for costs. These problems can be mitigated by making randomised evaluation design an integral part of the policy process, research design is not then an additional cost on top of policy implementation, instead implementation is designed in such a way that it allows for scientific evaluation after a period of time. The trials could be designed to contribute to learning as they went along.

Conclusions

This paper sets out a path for a research evaluation strategy. The overall argument is that civil renewal is a policy area that requires a deliberately experimental approach. It suggests that an initial exploratory phase of policy creation should be supported by research drawing on a number of approaches but in particular of the emerging concept of design experiments. In a second phase of work the focus moves to the more hard-nosed assessment of whether an intervention has made a positive impact. In particular emphasis is given to the potential of randomised control trials to deliver definitive results about 'what works'.

If a research and policy development strategy along the lines proposed in this paper is adopted it will make significant demands on the research community, practitioners and policy makers. The demands on the research community to become confident and competent in new research methodologies will be considerable. We also need researchers who are comfortable working alongside policy makers and practitioners. Such a capacity is central to the delivery of both design experiments and randomised controlled trials.

For practitioners there is a need to be actively involved in the search for what works rather than wait for it to be handed down on slabs of stone or perhaps more realistically by way of conferences, consultancies, handbooks and websites. Discovering effective practice is a challenge in which practitioners should be actively engaged. But more than that they should care about the 'what works' question and be supportive participants in research. This means getting involved beyond the heroic efforts made by many in completing survey forms. It means allowing the spirit of research to enter their daily work. More pragmatically, it means becoming part of investigative teams alongside researchers and policy makers.

As for policy makers the challenge is to use research not to confirm a pre-ordained direction of travel but as a genuine tool of investigation. The concept of piloting is given considerable lip-service but in the area of civil renewal it has to be a central and delivered

feature of policy development. Moreover the challenge is to provide a policy that can steer in defined directions and preclude some unacceptable options but allow scope for local initiative and experiment.

References

Bennett, T (1996) 'What's New in Evaluation research. A Note on the Pawson and Tilley Article.' *British Journal of Criminology*, 36(4), 567-73

Blunkett, D (2003a) *Civil Renewal : A New Agenda*. Edith Kahn Memorial Lecture. London: Home Office

Blunkett, D (2003b) *Active Citizens, Strong Communities*. *Progressing Civil Renewal.* London: Home Office

Campbell, D (1979) ' Reforms as Experiments' in J. Bynner and K. Stribley (ed) *Social Research: Principles and Procedures*. London: Longman

Clarke, A (1999) *Evaluation Research*. London: Sage

Chen, H (1990) *Theory-Driven Evaluations* . Newbury Park, CA: Sage

Cobb, P; Confrey, J; diSessa, A; Lehrer, R; Schauble, L (2003) 'Design Experiments in Educational Research ' Educational Researcher, 32(1), 9-13, January/ February

Design-Based Research Collective (DBRC) (2003) 'Design Based Research: An Emerging Paradigm for Educational Enquiry' *Educational Researcher,* 32 (1), 5-8, January/February.

Gerber, A and Green, D (2000) 'The Effects of Canvassing, Telephone Calls and Direct Mail on Voter Turnout: A Field Experiment ' *American Political Science Review* , 90(3), September, 653-63

Green, D and Gerber, A (2002) 'Reclaiming the Experimental Tradition in Political Science' in H. Milner and I.Katznelson (ed) *State of the Discipline*, Vol 111. New York: Norton, 805-832

Green, D; Gerber, A ; Nickerson, D (2003)' Getting out the Vote in Local Elections: Results from Six Door-to-Door Canvassing Experiments' *The Journal of Politics* , 65(4) , 1083-96

Greenberg, D; Linkz, D and Mandell, M (2003) *Social Experimentation and Public Policy Making.* Washington, DC: The Urban Institute Press

Hogwood, B (2000) 'The consideration of social experiments in the UK: Policy, Political, Ethical and Other Concerns' Paper to ESRC Conference on Social Experiments, 9-10 November

Jowell, R (2003) Trying it Out. The Role of 'Pilots' in Policy-Making. London: Cabinet Office

Lowndes, V; Pratchett, L and Stoker, G (2004) The CLEAR framework, mimeo. Available from Stoker

.

Magenta Book (2003) *Guidance Notes for Policy Evaluation and Analysis*. London: Cabinet Office

McCandliss, B; Kalcham, M; Bryant,P (2003) ' Design Experiments and Laboratory Approaches to Learning: Steps Towards Collaborative Exchange' *Educational Researcher.* 32(1), 14-16, January/February

McGraw, K (1996) ‘Political Methodology: Research Design and Experimental Methods’ in R.Goodin and H-D Klingemann (ed) *A New Handbook of Political Science*. Oxford: OUP

Medical Research Council (MRC) (2000) *A Framework for the Development and Evaluation of RCTs for Complex Interventions to Improve Health* (mimeo)

Medical Research Council (MRC) (2002) *Cluster randomised trials. Methodological and Ethical Considerations* (mimeo)

Moffit, R (2003) ‘The Role of Randomized field Trials in Social Science Research: A Perspective from Evaluations of Reforms of Social Welfare Programs’

National Research Council (NRC) (2002) *Scientific Research in Education*. Washington: National Academy of Sciences

Oakley, A (2001) *Evidence Informed Policy and Practice: Challenges for Social Science*. Manchester: Manchester Statistical Society

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

Pawson, R (2001) *Evidence-Based Policy: 11. The Promise of ‘Realist Synthesis’*. ESRC UK Centre for Evidence Based policy and Practice, Queen Mary, University of London

Shavelson, R; Philips, D; Towne, L; Feuer, M (2003) ‘ On the Science of Education Design Studies’ *Educational Researcher*, 32(1), 25-8, January/February

Appendix: Scientific Principles

The six principles identified below borrow and paraphrase material from NRC, 2002, 3-5 and the wider discussion in Chapter 3 of the NRC report.

SCIENTIFIC PRINCIPLE 1

Pose significant questions that can be investigated empirically

Specifying a worthwhile question is essential to scientific research. The question needs to be posed in a way that enables the testing of various alternative answers. Ultimately, the final court of appeal for the viability of a scientific hypothesis or conjecture is its empirical adequacy. The questions and the research programme developed to address them must also reflect a solid understanding of the relevant theoretical, methodological, and empirical work that has come before.

SCIENTIFIC PRINCIPLE 2

Link research to relevant theory

It is the long-term goal of much of science to generate theories that can offer stable explanations for phenomena that stretch beyond the particular. Social science, it is true, tends to reach for more middle-range theories that can account for some aspects of the social world in some conditions rather than grand general theories. Crucially every scientific inquiry should be linked, either implicitly or explicitly, to some theory or conceptual framework that guides the entire investigation.

SCIENTIFIC PRINCIPLE 3

Use methods that permit direct investigation of the question

Methods can only be judged in terms of their appropriateness and effectiveness in addressing a particular research question. Most scientific claims are significantly strengthened when they are subject to testing by multiple methods. What methods to choose can depend not only on the specific research question but also the stage of the research process.

SCIENTIFIC PRINCIPLE 4

Provide a coherent and explicit chain of reasoning

Inferential science is at the core of science: explanations, conclusions, or predictions based on what is known and observed. The aim is to establish connections between phenomena that could meet the strictures of a sceptical observer. The validity of inferences made through this process is strengthened by identifying limitations and biases, and estimating uncertainty and error. It also involves ruling out plausible counter explanations in a rational, compelling way. Detailed descriptions of procedures and analyses are critical to permit others to critique, to analyse, and to attempt to replicate, a study.

SCIENTIFIC PRINCIPLE 5

Replicate and generalize across studies

Scientific inquiry emphasises checking and validating individual findings and results. Since all studies rely on a limited set of observations, a key question is how individual findings generalise to broader populations and settings. Ultimately, scientific knowledge advances when findings are reproduced in a range of times and places and when findings are integrated and synthesized. The argument for a team work approach to research questions is based on the need to develop a range of connected studies in order to boost the scope of scientific validity.

SCIENTIFIC PRINCIPLE 6

Disclose research to encourage professional scrutiny and critique

Science is a community endeavour. It relies on results being shared and subjected to professional scrutiny by peers. This option of public critique is an indication of the health of a scientific enterprise. Indeed, the objectivity of science derives from publicly enforced norms of the professional community of scientists, rather than from the character traits of any individual person or design features of any study.