

Research strategies for civil renewal

Gerry Stoker and Stephen Greasley
IPEG, University of Manchester

Contents

1. Introduction
2. Civil renewal as a policy area
3. Setting the scientific standard
4. The contribution of best practice case studies: strengths and limitations
5. Tools for exploratory evaluation: introducing design experiments
6. Tools for impact evaluation: the scope for randomised controlled trials
7. Conclusion: the challenge of interdependence for policy makers, practitioners and researchers
8. References

Appendix: How the tools work in practice

1. Introduction

The civil renewal agenda is one that attracts a great deal of interest from policy makers. They want to know what is changing in communities, what is happening to citizenship and what interventions will work. They want to know 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?

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. It has to rely on a tight-loose formulation (Mulgan and Lee, 2001), where it sets broad goals and objectives but must involve others with local knowledge at a minimum in implementation and design processes. Civil renewal, by definition, cannot be achieved except by involving and getting the commitment of others.

Because of these factors, 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.

Developing policy through experiments is likely to be a challenge to policy makers and practitioners; a point considered further in the conclusion of this paper. Certainly working in this way is going to be a challenge to social researchers in their role in evaluating policy options and developments. The questions for social scientists interested in connecting research to policy are how they can best advise and how can their advice best be warranted? In short, how can they demonstrate a basis for the knowledge claims that they make? How can researchers help identify 'what works'?

The key to making progress, according to this paper, is not some magic new method. There has been rather too much of a 'methods war' in evaluation work. The perspective of this paper is informed by a commitment to utmost rigour in social science research combined with an understanding of policy and implementation processes particularly in respect of civil renewal. A variety of methods can then be seen as appropriate to a process of discovery. The aim will not be to discuss what should be studied, but rather how it should be studied.

There is a considerable body of practice and reflection on evaluation from which we can draw to develop an approach suitable to the area of civil renewal. Education, health, crime and other social science research has in the UK, and especially in the US, seen intensive debates about what are appropriate social scientific approaches to understanding complex social issues and intervening effectively to achieve positive outcomes. The main thrust of this document is to suggest that it will be necessary to develop a framework for evaluation that moves from an exploratory formative phase towards a definitive and long-term judgement about the impact of an intervention.

The paper begins with a review of civil renewal as a policy area because it is through that understanding that an appropriate social science response can best be identified. The nature of the policy area and the substance of the questions at stake make a difference in terms of selecting methods. The scientific standard that should be the focus of production of warrantable and authoritative knowledge from researchers is the subject matter of the next section. We need guidelines to maximise the prospect that research results will be valid and rigorously supported by evidence. These preliminaries set the framework for the research response.

The report then examines the contribution made by the identification of best practice. Several early policy papers on civil renewal called for the identification of best practice (Blunkett, 2003 a and b). For the purposes of this paper, best practice is defined as effective practice and is seen as about evidence produced for the manager/implementer; it is evidence about how to do things. As such it is an essential part of policy development in civil renewal as in many other areas. However, in so far as its identification rests on a series of case studies that may not be examined and tested in a systematic way and are not replicated experiments, it cannot establish what policy works. Effective practice reports can and should provide a detailed descriptive account that enables those in the field to be motivated and

inspired to design initiatives to meet the circumstances they are in. The paper makes some suggestions about how effective practice can best be identified and propagated. What is clear, however, is that to establish what works requires a more systematic process of search and reflection.

To meet the challenge of finding out what works we need actively to examine the policy options that are available and the casual connections between interventions and outcomes. In Section 5 of the report it is argued that there should often be an exploratory phase to this work in respect of civil renewal. The methods suitable to this formative phase of evaluation are explored in that section. The emphasis should be on exchange and feedback among researchers, practitioners and policy makers that are trying to achieve positive change. The report notes the value of systematically reviewing earlier studies and modelling the key factors that might drive civil renewal. It promotes the idea of design experiments as an interesting and active way of undertaking exploratory investigation.

Ultimately the goal of evaluation should be to show that some interventions have made a positive difference greater than that which could be achieved by alternatives. To answer the 'what works' question means ultimately being able to demonstrate the impact of an intervention, what effect was achieved and how it was achieved. Again a variety of evaluation methods are available and particular attention is paid to the prospects of an approach to establish cause and effect using randomised controlled trials. There is no agenda to develop an exclusive advocacy of these methods; it is simply argued in this paper that these methods are particularly appropriate to the civil renewal policy agenda and that their adoption maximises the chances of meeting the challenge of producing valid and authoritative social science knowledge.

The argument presented here is as much a challenge to policy makers and practitioners as it is to researchers. To that extent it marks a break from the strategy of realistic evaluation advocated in Pawson and Tilley (1997). 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, 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 or, worse, impressions of limited value. The theories of change 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.

2. The nature of civil renewal as a policy area

Civil renewal is about giving people a stronger sense of involvement in their communities and a greater say over their lives. It is about encouraging people to be active citizens, strengthening communities and enabling partnership between citizens and officialdom in the planning and delivery of public services (Blunkett, 2003b: 6). It involves a large range of potential interventions: examples from Home Office policy alone include:

- Building communities' capacity to engage in the governance of public services through asset transfer, community development and citizenship education
- Engaging local communities in shaping and delivering the integration of refugees
- Involving more citizens in the administration of restorative justice to improve victims' satisfaction, support conflict resolution, and reduce re-offending
- Strengthening community and neighbourhood level influence on policing through the police reform programme
- Improving the community engagement dimension in the work of the Drug Action Teams and the Crime and Disorder Reduction Partnerships
- Ensuring communities are involved in shaping the outcomes of the Anti-social Behaviour Trailblazers
- Connecting and maximizing the community engagement impact of the police public reassurance and CJS confidence work

- Developing community cohesion through involving citizens from all backgrounds to break down barriers
- Involving faith communities in shaping public policies and services
- Encouraging local authorities to lead on civil renewal through the development of ‘civic pioneers’.

There are also a wide range of potential collaborations in policy development with other government departments (neighbourhood governance with ODPM; cultural activities with DCMS; tackling rural exclusion with DEFRA; public involvement with DoH; community involvement and citizen development in schools with DfES).

The discussion above does not do justice to the depth and variety of the interventions that could be developed under the heading of civil renewal. It is aimed at giving sufficient flavour of the civil renewal agenda to establish three points. First civil renewal is a policy area that cannot be developed in a simple top-down manner.

As Mulgan and Lee (2001) explain, there are a few simple policy areas where the process of delivery can follow a fairly straightforward model: objectives are defined by central government elected politicians, policy to achieve objectives is designed by civil servants, implementation is handed over to a delivery agency, and policy is delivered. Civil renewal certainly does not fall into that model. First it is not clear what is the right mechanism or mix of mechanisms to deliver the objective, the resources to implement any change are not under the control of central government alone, change may require new cultural attitudes and behaviours and the whole process is complicated by the over-lapping nature of a range of potential interventions. In short, civil renewal is bound to take the form of a ‘tight-loose’ form of policy development where the centre can set broad objectives but the process of policy development will require adaptation, experimentation and development in the light of commitment and capacity on the ground. The government is not going to be in a position to advocate specific reforms, knowing for certain they will be successful. They are involved inevitably in a process of exploration. In the light of this understanding the role of research evaluation can be seen as central to the process of policy development.

The second point that flows from the nature of the civil renewal agenda is that interventions associated with that agenda are likely to constitute a series of ‘complex interventions’.

According to the Medical Research Council (2000), complex interventions have two qualities. Firstly, the intervention comprises a range of elements, all of which are essential to its proper functioning. Second, what exactly are the active ingredients and how they relate to one another is difficult to establish. Most interventions in the arena of civil renewal are likely to share those qualities. Is an intervention in a community driven to success by the institutional form, its particular content, the personalities of the people involved, the social status relationship of the deliverers and receivers, organisational culture and leadership, the time scale of the project, or by some mix of all these and maybe other ingredients? Again, the case for a systematic research evaluation approach to help unravel this complexity seems clear and undeniable.

The third characteristic of civil renewal interventions that adds to and helps to define the research challenge is that interventions are likely to be aimed at communities rather than isolated individuals. A trial of a new drug focuses on the impact on individuals but a civil renewal intervention, although it works through individuals, is likely to be concerned with an impact in a community. There are particular research challenges in trying to understand community-level effects and these will need to be taken into account in developing research programmes.

3. Setting the scientific standard

Social scientists do not agree all the time and they certainly do not agree over how best to engage in evaluative research. Notwithstanding these differences, it might be possible to provide a set of scientific principles on which many researchers would agree. As social scientists we would all value evidence, scepticism about claims to knowledge, a keenness to rule out alternative explanations and a commitment to constructive criticism as key ingredients in developing knowledge. Such an ethos does not automatically promote particular methods or approaches to understanding social and political phenomenon, but it could provide guidance to encourage the most rigorous and systematic approach to the collection of evidence in respect of the development of public policy. As the Government's Chief Social Researcher (Davies, 2004, 3) notes, 'not all research is of sufficient quality to form the basis of sound policy making'. Research may be undermined by lack of clear objectives, weak methodology, inadequate reporting and analysis, selective use of the data

and as a result may come to conclusions that are not justified. The aim of the principles set out below is to establish an ambition to improve the quality of research in respect of the civil renewal agenda.

An investigation by the Education Committee of the National Research Council of the USA provides a helpful starting point. It had the task of considering how to raise the quality of applied education research and suggested that a key way forward was for all researchers to work against a template of scientific endeavour. The Committee (NRC, 2002, 2) argued:

At its core, scientific inquiry is the same in all fields. Scientific research, whether in education, physics, anthropology, molecular biology, or economics, is a continual process of rigorous reasoning supported by a dynamic interplay among methods, theories, and findings. It builds understandings in the form of models or theories that can be tested. Advances in scientific knowledge are achieved by the self-regulating norms of the scientific community over time, not, as sometimes believed, by the mechanistic application of a particular scientific method to a static set of questions.

The process of accumulating knowledge is a complex one, with steps forward and backward. Indeed challenge is central to the process. Research and lesson-drawing takes time and the undertaking of a range of studies. To develop core knowledge in an area will require a variety of methods and it will need a significant body of researchers. Generally the argument of the Committee lends its weight against the lone scholar model of research and argues instead for integrated teams sharing research objectives and challenging each other's performance and insights.

What is essential is that researchers have a set of guidelines to steer their activity. As the Committee explains (NRC, 2002, 2):

These principles are not a set of rigid standards for conducting and evaluating individual studies, but rather are a set of norms enforced by the community of researchers that shape scientific understanding.

The six principles identified by the Committee are presented below in a slightly adapted and summarised form (the discussion below borrows and paraphrases 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.

The scientific principles set out by the National Research Council of the United States provide a very useful framework for setting the template against which our endeavour to understand the processes of civil renewal needs to be judged. The application of the principles will vary from setting to setting. However they do provide an essential guideline in the review of research evaluation approaches undertaken in this paper.

4. The contribution of best practice case studies: strengths and limitations

One of the characteristics of civil renewal as a policy area is that there is plenty of innovation, experimentation and activity from which it should be possible to learn lessons. There remains, however, an initial challenge, which is to be able to collect an effective account of

what is being done. Then there is the issue of the value of the information that is collected and how it is to be communicated and to what end. We need to consider both issues.

The challenge of identifying effective practice is a considerable one and will require a strategy based on systematic search. There are several efforts that have been undertaken or that are underway to understand more and learn more from effective practice. The scale and nature of the challenge is similar to that in the area of health improvement. In that context, a programme is underway to develop a 'learning from effective practice system' (French, 2004). The strategy of the Health Development Agency (HDA) rests on a recognition that 'very little local practice is written up, captured or shared in a systematic way amongst practitioners or with wider communities of practice or policy makers'. A few cases are written up but there is a sense in which only the surface of practice out there is being touched. This argument made in the context of health policy might well apply in the area of civil renewal.

Given the scale of civil renewal activity the challenge of developing an effective practice strategy needs to be developed along lines similar to that suggested for health improvement by the HDA. There needs to be a systematic search, drawing in practitioners as well as researchers, to establish some examples of effective practice which can be defined as 'action that can demonstrably be shown to have created positive change (French, 2004, 5). One option would be to develop some national guidelines for what constitutes effective practice and some minimum standards about what information needs to be collected, but it is clearly necessary to recognise the difficulties of demonstrating strong causal links using case studies and not to over-burden practitioners in terms of information requirements.

What researching effective practice does is provide crucial 'how to' information for practitioners. An individual case cannot establish 'what works' but it can help a practical administrator think through the dynamics of implementation and consider how a particular practice can be transferred from one setting to another. What is required is a detailed enough narrative of how the scheme works to enable those insights to be gleaned. The review of effective practice collection from the HDA suggests two further general lessons: one, the material needs to be kept up to date and marketed in a way that ensures knowledge of its existence and ease of access; two, the material needs to be collected with the active co-operation of practitioners. Without that co-operation ownership may be reduced and take-up of the examples of effective practice diminished.

The core value of effective practice collecting, by way of case studies, is as an implementation aid. It would be a mistake, however, to assume that a collection of disparate case studies can deliver a basis for a judgement of 'what works' that meets the scientific standards set out in section 3. 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 generalisation (see NRC, 2002, 74).

5. Exploratory policy development: introducing design experiments

The idea of a framework for evaluation stretching from an initial exploratory phase to an attempt to come to a more definitive judgement about the impact of interventions underlies the next two sections of the paper. The variety of terminology and potential distinctions in the evaluation literature are quite voluminous (Clarke, 1999, Ch 1; the Magenta Book, 2003, Ch 1) so 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 standards set out in Section 3 of this paper. We need approaches that can help explore impact but also help us understand why change has occurred.

5.1 Systematic reviews

An obvious starting point is to review what has been done and what is already underway. The search for more effective practice outlined in the previous section can provide much initially useful material and 'food for thought'. The challenge to researchers is to sift through that material in a systematic manner. The information generated by such an exercise is likely to be full of uncertainties and contradictions. The world of civil renewal as in so many other

areas of public policy is one of uncontrolled experimentation and unsystematic research. The findings that are produced are often not cumulative and sometimes confusing. The answer recommended from several quarters is to engage in a systematic review (Oakley, 2001; Magenta Book, 2003, Ch 2; Pawson, 2001).

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. The search for effective practice, outlined in the previous section, may contribute to this process. 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. Without such clarification it is not possible to judge whether the systematic review has been subject to hidden or latent biases. Being explicit about the basis of the comparison enables others to judge more effectively the warrantable nature of what the evidence is being claimed to suggest.

5.2 Modelling and theories of change

The Medical Research Council (2000), in outlining its framework for evaluating complex interventions, makes considerable play of the attraction of theory building and modelling in early phases of the work. The aim is to identify the underlying mechanisms associated with an intervention in order to predict how they will influence outcomes. 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. At this exploratory stage these tentative theories of change can be tested in a variety of relatively low cost ways. 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.

5.3 The role of pilots

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.

As Jowell (2003) notes in a report for the Cabinet Office, a pilot could be a vital first stage in a longer-term evaluation. But it is essential that 'a pilot should be undertaken in a spirit of experimentation. So, if it is clear at the outset that a new policy and its delivery mechanisms are effectively cast in stone, a pilot is redundant and ought not to be undertaken' (Jowell, 2003, 5) . What is also crucial is that the pilot is given time to deliver results, the independence to undertake its work and present its findings, and that expectations of success, or for that matter failure are managed by avoiding tags such as 'trailblazer' that may be misinterpreted. With a commitment to rigorous methods and time to try out things in practice a pilot can deliver a neutral evaluation, which provides a genuine test of a policy.

Jowell (2003) emphasises the value of 'policy development' pilots. These pilots make the impact at the early stages of the policy process and could focus on the impact of a proposed intervention or explore the process or delivery mechanisms associated with an intervention, or in some cases both. The method most favoured by Jowell is randomised controlled trials (RCTs), although the value of other methods is also recognised. There are a number of strengths and weaknesses associated with RCTs that are examined further in the next section of the paper. The dilemma that needs to be explored at this stage of the discussion is how suitable RCTs are to the early stages of policy development. Crucially, Jowell (2003, 31)

notes the 'frequent conflicts between the demands of the policy cycle on the one hand and rigorous evaluation on the other'. The political pressure can be to roll out an intervention because of its perceived benefits and as a result the evaluators may find that the time to make a full impact analysis is squeezed. To find the way forward, some sort of compromise has to be struck between 'knowing everything' and 'knowing something'.

In the case of civil renewal a full RCT may not be the best initial option in part because the appropriate mechanisms for intervention may be so uncertain or unclear. An RCT that requires an intervention to be followed through without adaptation or change so that its impact can be rigorously assessed may not be the right initial step in empirical evaluation. It may be that a more explicitly exploratory and iterative phase would be value in the fist instance. This option is explored in the next section.

5.4 Design experiments as a form of piloting

In the exploratory phase on work an important option to consider is 'design experimentation' or 'design based research'. This strategy 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.

5.4.1 What are design experiments and why would we need them?

In educational research there are five characteristics of design experiments (DBRC, 2003, 5):

1. Design experiments have the goals of designing learning environments and developing theories of learning
2. The research proceeds though a cycle of design, enactment, analysis and redesign
3. The research leads to theories that can be communicated across contexts
4. The research studies how designs for educational learning function in real world contexts
5. The research uses methods that can connect the process of design and intervention with the needs of policy makers and practitioners

The first point presents the dual goals of design experimentation – the rejection or validation of an underlying theory and the detailed design and testing of an intervention based on that theory. These goals are achieved simultaneously as the intervention is implemented in a classroom. The second point describes how this process occurs. An intervention 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.

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 intervention such as a curriculum or teaching strategy. Points three and four are about the relationship between theory and design. The design of an intervention is based on clear theoretical presuppositions which are placed under scrutiny as the intervention proceeds. In the educational context the theory on which design is based is often concerned with cognitive and learning processes.

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 fifth point reinforces the idea that 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.

5.4.2 Applications of design experiments

Design experimentation has primarily been applied to education research. The stages that a design experiment would take are outlined below:

- Theory based design of intervention – this may focus directly on a teaching strategy, on training for teachers, on institutional contexts or often on multiple levels at the same time.
- Systematic data is collected prior to, during and after the programme. This will include measurement of outcomes of interest, but also collection of data on the processes of learning often including video taping and qualitative research.
- The collected data is analysed and adjustments to the intervention are suggested and the process is repeated.

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 recognises 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.

Design experiments may be capable of application in the civil renewal setting. Take for example the development of community forums: one body of theory would suggest that people often initially become involved in community issues to protect their own interests but under certain circumstance may develop a more rounded view of the common good as they debate issues with others who express different views. The initial activity can then develop into further civic engagement. A design experiment, in the light of this proposition, would try to find the best way to ensure that a community forum is run in a way that draws citizens' attentions to alternative views and leads them to think about the common good. This may involve trying to discover the right balance between time committed to information presentation and time committed to debate, how the information should be presented, how to ensure that different elements of the community are represented. It could involve identifying the right mix of professional or citizen leadership and ways of structuring meetings that allows controversial issues to be addressed without slipping into mutual antagonism.

Institutional reform is a second example where design experiments might be used. The underlying theory would refer to the importance of public sector organisations actively encouraging involvement from citizens and the impact that can have on general citizen engagement. A design experiment might then involve varying the forms of institutional support for citizen engagement, to see which approach is more likely to achieve higher participation, and whether the response of citizens varies between groups and social contexts. This systematic knowledge can then be fed into the development and implementation of policy.

5.4.3 Judging the contribution of design experiments

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 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.

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.

6. Tools for impact evaluation: the scope of randomised controlled trials

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). The Medical Research Council’s framework for evaluation moving from formative phases to the long-term makes a distinction between exploratory RCTs and definitive RCTs. The former may be appropriate tools in the piloting phase (as discussed in **5.3**) and as such some of the normal rules of RCTs may be relaxed with scope given to vary and adjust the intervention. The aim of a definitive test is to judge the impact of the intervention. At that 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. RCTs are not the only way of engaging in long-term evaluation and the scope for other methods is considered in the second half of this section of the report.

6.1 Randomised Controlled Trials: what is involved?

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.

6.2 Possible applications of RCTs to civil renewal

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).

6.3 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 as outlined in Section 3.

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'

6.4 The need to compliment RCTs with other evaluation tools

There are other evaluation options to be considered as part of an overall framework for evaluation. It is not necessary to dwell in depth on these options as they are dealt with in detail in a range of reports and sources (Magenta Book, 2003; ESRC UK Centre for Evidence Based Policy and Practice, Cabinet Office Policy Hub). There is a wide variety of quasi-experimental and other after-the-event evaluation options that are available stretching from longitudinal studies through economic appraisal to various form of qualitative analysis. The evaluation of the introduction of citizenship into the curriculum provides a good model of the way after-the-event evaluations that can contribute to a judgement about the overall impact of an initiative and at the same time provide lessons and insights as the research moves forward (Kerr et al, 2004). As more evidence is collected there will eventually be an argument for a number of meta-analyses that sift and then bring together the cumulative results of a range of studies.

7. Conclusion: the challenge of interdependence for policy makers, practitioners and researchers

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 exploratory approach. It is a policy area that should be guided by a spirit of search and investigation because the aims of civil renewal and the interventions that will achieve those goals are not immediately obvious. It is a policy that demands a spirit of partnership as well because it is not an area where the will of central government to see change is enough to guarantee that there will be change.

Having made an argument for a spirit of exploration, however, the paper does not suggest letting 'a hundred flowers bloom' without guidelines and guidance. Rather it argues that policy needs to be supported in its development by research that achieves the highest scientific qualities. 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. Again, a range of methodologies are recognised as having an important contribution, but 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. As the US National Research Council report on research in education (NRC, 2002, 63) puts it, the challenge is disarmingly simple and yet incredibly difficult:

Debates about method - in many disciplines and fields - have raged for centuries as researchers have battled over the relative merit of the various techniques of their trade. The simple truth is that the method used to conduct scientific research must fit the question posed, and the investigator must competently implement the method.

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

Davies, P (2004) *Is Evidence-Based Government Possible?* Jerry Lee Lecture. 4th Annual Campbell Collaboration Colloquium, Washington DC, 19 February

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

French, J (2004) *The Learning From Effective Practice System*. Health Development Agency

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

Kerr, D (2004) *Making Citizenship Real*. Second Annual Report of the nine-year Citizenship Education Longitudinal Study, conducted by NFER on behalf of the DfES.

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'

Mulgan, G and Lee, A (2001) *Better Policy Delivery and Design. A Discussion Paper*. London: Cabinet Office

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

How the tools work in practice

1. Introduction

This appendix illustrates how the approach to research outlined in the general document might work in a specific case. As was made clear in the main paper, a key feature of the research approach advocated is that it develops alongside policy rather than as after-thought. So it is necessary to think of an illustrative policy that would appear to be broadly in line with civil renewal thinking. For the Home Secretary, David Blunkett, one core goal of civil renewal is clear:

'My objective is nothing less than changing the institutional structures and cultures of our public sector by enhancing the degree and quality of citizens' involvement in shaping policies and services affecting their lives'¹

One policy that might be seen as delivering on that objective is the emergence of community-based housing organisations for the control of publicly-owned housing stock. Rather than the experience of management through large-scale bureaucracy run through the council, a management trust or a housing association the aim of the policy would be the development of community-owned and managed housing estates. Such a policy was pioneered in Scotland in the mid-1980s and was the focus of research both at its launch² and in a re-visit evaluation in the mid-1990s³ For the purposes of the illustration this prior research is helpful in indicating key factors behind the success and limitations of the schemes. For the illustration it is assumed that there is a broad policy commitment from ministers to develop community-based ownership of public housing as they regard that form of management to be likely to deliver a more responsive service and provide a wider base for civil renewal. The aim of the research programme is to test those assumptions and aid in the design and development of the policy.

The illustration of a research approach presented here picks up on two key themes of the main paper. The first is the need for the development of a research programme to take up the challenge of the scientific principles laid out in Section 3 and to develop a programme of work that runs in parallel with an initial developmental phase in the policy process through to large-scale implementation.

2. Thinking of good questions: connecting to what we know

The importance of asking the right questions is central to any scientific investigation. The questions must be significant and they must be posed in such a way that it is possible to test the adequacy of alternative answers. The overarching question in the case of the policy under investigation in this appendix is does a community-ownership approach in housing management have advantages in service responsiveness and civil renewal? A more specific

¹ The Politics of Civil Renewal, notes for talk to be given by D. Blunkett in The United States

² D. Clapham and K. Kintrea 'Community Ownership and the Break-up of Council Housing in Britain' *Journal of Social Policy*, 23, 1994

³ Clapham, D., Kintrea, K. and Kay, H. (2000) 'User Participation in Community Housing: Is Small Really Beautiful?', Stoker, G. (ed) *The New Politics of Local Governance*, London : MacMillan

set of questions is likely to emerge given a more detailed consideration of the proposal under investigation.

One important way of developing the more detailed questions to be investigated is to draw on what we know. In this particular case we can draw on some general theories of what drives service responsiveness and what drives civic participation. There is also more detailed and specific evidence. As the guiding scientific principle 2 stresses, progress and understanding are achieved by linking theoretical and empirical work.

Tables 1 and 2 provide illustrations of a synthesis of key insights from a range of past studies looking respectively at how to improve service responsiveness and engage citizens in decision-making.⁴

Table 1: Change to achieve service responsiveness

Aspect	Transformational change	Continuous improvement
General approach	Starts with a clear high level vision of desired future state but no clear plan of exactly how it might be achieved. Identifies major opportunities in line with vision and uses those to lever change. Can result in public authority doing different things in different ways, or supporting others in these tasks. Challenges purpose and rethinks the client and supply mix. High level strategic partnerships often important.	Basic philosophy is to start with what is there and work out how to make it better. Performance management through the public authority drives a focus on how to deliver services better (which can involve major redesign of approach and re-engineering of service processes). Best when involves frontline staff in process of service improvement.
Change	Experimental, discontinuous,	Focused, consistent effort to

⁴ These tables and the arguments reflected in them do not necessarily reflect the fullest or most appropriate literature review but for the purpose of the illustration they indicate the sort of processes and thought processes the policy or research programme would need to address.

pattern	fluid, uncertain, organic. Learning and adaptive behaviour critical. Can feel chaotic - that everything is changing at once.	raise performance. Organisation remains stable as change is managed through different priority areas.
Management style	High level leadership is the key. Onus on creativity and partnership building.	Leadership distributed through organisation. Onus on management skills.
Risk and sustainability	High risk, potentially high reward. May be difficult to sustain unless programme of embedding established	Reduced risk, but danger of initiatives running out of steam or taking cautious route

Source: adapted from A. Melville, What Drives Service Improvement? IdeA 2002

Table 1 suggests that the policy that is being proposed implies a radical shift to community management and therefore might be closer in dynamic to transformational change. Table 2 suggests factors that may have to be present or enhanced in order to encourage citizens to take on the more involved roles required if a community ownership scheme is to be effective.

Table 2: Factors promoting participation: it's CLEAR

Factor affecting participation	How it works	Associated policy target
Can do	The individual resources that people have to mobilise and organise (speaking, writing and technical skills, as well as confidence to use them) make a difference in their capacity to participate	Capacity building
Like to	To commit to participation requires a	Sense of community, civic engagement, social

	sense of involvement with the public entity that is the focus of engagement	capital and citizenship
Enabled to	The civic infrastructure of groups and umbrella organisations makes a difference because it creates or blocks an opportunity structure for participation	To build the civic infrastructure so that there are groups and organisations around to channel and facilitate participation
Asked to	Mobilising people into participation by asking for their input can make a big difference	Mobilisation schemes that are diverse and reflexive
Responded to	When asked people say they will participate if they are listened to, not necessarily agreed with, but able to see a response	A public policy system and institutional environment that can show a capacity to respond

Source: adapted from G. Stoker (in collaboration with V. Lowndes and L. Pratchett)contribution to ESRC policy seminar on civil renewal, September 2003

In terms of the development and implementation of the policy Table 2 suggests that capacity building, community cohesion, support for communities, an active mobilisation strategy and a responsive environment of public bodies might be key factors in determining the success or otherwise of the scheme.

To develop more specific questions about the policy it is necessary to examine the existing evidence of similar schemes in more detail. Here the research work conducted by Clapham and Kintrea referred to earlier comes into its own. 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. In 1989 a survey of four of the first CBHOs, including the most deprived, showed that the majority of tenants were significantly more satisfied with CBHO than they had been with council housing services, 68% across all schemes claiming that the overall service was better. By 1994, with the exception of one project, the general consensus was that CBHO control of housing remained superior, 81% of tenants being satisfied, a 9% increase from 1989. Indeed, 40% of interviewees from all four schemes believed that the service was improving and only 6% stated it was getting worse.

Measures of management effectiveness suggested that after a year CBHOs had instigated better housing management without increasing repair and management expenses. In the majority of the areas studied, concerns over neighbourhood problems decreased after the CBHOs were established. In the areas where it increased it may have been a result of residents transferring their concerns because their housing concerns had been addressed. Although the distribution of most popular and unpopular areas remained unchanged, measurements of location satisfaction improved in the CBHOs after they were first set up. All CBHO residents reported a drop in their levels of worry about crime but two CBHOs showed an increased concern about safety after nightfall.

Community-based housing organisations were regarded more highly than the council, only 44% of interviewees believing that the council can be 'trusted to do what is right' as opposed to 75% believing that of the CBHOs. Levels of attendance at AGMs were high and the

leading management of CBHOs - drawn from about 15 per cent of the resident population - was shown to comprise of the average or near average resident in terms of social and educational characteristics.

The biggest internal constraint on the participative culture of CBHOs was that staff employed by the CBHOs used their technical knowledge and full-time position to set agenda for residents and their influence decision-making. The greatest external barrier to participation was the dependence of CBHOs on other institutions for their funding and broader activities. What the residents wanted and wanted to do was constrained by powerful institutions around them.

Depending on the policy under investigation there will undoubtedly be some research evidence and theoretical work (how much will vary) that can address issues relevant to the policy either in general terms or more specifically. In some cases this material gathered in a process of review will be sufficient to move on to the design of a final policy programme and its assessment. In other cases however, it might be helpful to think of having one more step to explore the policy options and develop better measurements of the outcomes that are being targeted.

3. Design experiments to develop and test the policy (and scientific understanding and measurement tools)

The National Research Council's (NRC, 2002) discussion of the third scientific guiding principle which states that methods should permit direct investigation of the question under review is especially challenging. The argument presented here is that in some cases design experiments, as outlined in Section 5.4, might help meet that challenge because they would enable a further systematic sifting of the key variables that might affect the outcome of any reform. Moreover they would expose clearly the strengths but also the limitation of any measurement undertaken. Measurement techniques and instruments might be improved in the period of the design experiment and a clearer estimate of the degree of uncertainty surrounding any results would emerge. In short the design experiment phase might advance both the understanding of the policy and the scientific challenge involved in impact evaluation.

The research conducted on CBHOs and the general level of knowledge about service improvement and the encouragement of participation give a good basis for developing policy in this field. 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.

A design experiment would pilot 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 they would have been without the design experiments.

For a concrete example, one concern might be how to design CBHOs and other institutions in the community to improve the chances that active civic engagement will persist. Starting from the CLEAR criteria design experiments would systematically implement, monitor and vary policy interventions. Table 3, below, highlights how some specific and practical concerns of institutional and policy design are linked to the CLEAR theoretical framework. This is of course a very extensive research programme, but the value of a design experiment is that it is both practical and theoretical. The 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 3: Promoting participation: from theory to design

Participation factors	Associated policy target	Design experiment
Can do	Capacity building	<p>Research questions: what is the best way to provide the skills needed by citizens? Would these techniques vary between social groups?</p> <p>Possible areas of investigation:</p> <ul style="list-style-type: none"> • Learning by doing • Formal courses • Literature • Interactive computer based learning • Peer-group led learning
Like to	Sense of community, civic engagement, social capital and citizenship	<p>Research questions: how can public action enhance sense of community?</p> <p>Possible areas of investigation:</p> <ul style="list-style-type: none"> • 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	To build the civic infrastructure so that there are groups and organisations around to channel and facilitate participation	<p>Research questions: can membership of social/leisure groups lead to more civic engagement?</p> <p>Possible areas of investigation:</p> <ul style="list-style-type: none"> • 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	Mobilisation schemes that are diverse and reflexive	<p>Research questions: what is the best way to disseminate information about opportunities for activity?</p> <p>Possible areas of investigation:</p> <ul style="list-style-type: none"> • 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	A public policy system and institutional environment that can show a capacity to respond	<p>Research question: what features of institutional design affect the likelihood of citizens becoming involved?</p> <p>Possible areas of investigation:</p> <ul style="list-style-type: none"> • 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

4. Reflecting on the design experiments: the development of coherent policy and research

As the NRC (2002) guidelines suggest in identifying their fourth principle, the key to good science is good measurement and then the explicit linking of a chain of reasoning that systematically links empirical observation to underlying theory. After the initial search of the available evidence and theory, perhaps supplemented by insights from design experiment, there is a need for a period of systematic investigation. That investigation would require researchers to specify the form and conditions under which CBHOs would deliver the best outcomes, and policy makers and practitioners to judge the extent to which they want to commit to reform using CBHOs.

5. Replication in the form of randomised experiments trials

By this stage the policy has been refined, 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. As the NRC (2002) report argues science requires replication. In the social sciences this task can be very difficult to achieve because differences in context (setting and time) make replication problematic. As the report goes on to suggest, what is required is 'a more nuanced notion of replication' (NRC, 2002, 71). A range of statistical and non-statistical mechanisms are available to enable generalisations to be tested but in many cases 'subsequent applications, implementations or trials are often necessary to assure **generalizability** or to clarify its limits' (NRC, 2002, 71).

Randomised controlled trials (RCTs) as discussed in Section 6.1 provide a powerful and uniquely valuable tool in searching for knowledge that can be seen as applicable in a range of settings and beyond the experience of a specific case. In the case under discussion the refined CBHO intervention could be contrasted with sites of non-intervention. The challenges of design RCTs are considerable as noted in Section 6.1 but they are not insurmountable.

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 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 (see section 6.1). 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.

6. Disclose and review: openness among researchers, policy makers and practitioners.

The final scientific principle identified by the NRC (2002) rests on a commitment among scientists to share data and to open up their investigations to others. This principle has a wider application on the context of the development of civil renewal strategies. The research strategy outlined in the main body of this report rested on two claims. One was that appropriate policy in this area is uncertain and the other was that interventions to be

successful by definition involve a positive response by more than government alone. For these reasons it was argued that research needed to run alongside reform. For that to occur, an openness in exchanges between policy makers, practitioners and researchers throughout the process from policy initiation through to its delivery and ultimate impact evaluation is required.