Design Experiments: a Programme for Policy Learning and Evaluation

Peter John and Gerry Stoker, Institute for Political and Economic Governance, School of Social Sciences, University of Manchester, Oxford Road, Manchester, M13 9PL, UK E-mail: peter.john@manchester.ac.uk, gerry.stoker@manchester.ac.uk

Design Experiments: a Programme for Policy Learning and Evaluation

Abstract

This paper sets out the rationale and method of design experiments, which are an intensive form of research on innovations in the public sector. Design experiments rely on a close collaboration between researchers and practitioners, who both implement the intervention and observe it. The method originally derives from the design sciences, such as aeronautics and artificial intelligence, where designers use the information about the implementation of an intervention to help re-specify and re-calibrate it until it works. The paper reviews the existing literature on design experiments, which is based on educational practice; defines the necessary conditions for a design experiment; proposes some modifications to the educational design model to help it apply to conditions outside the classroom; and compares and contrasts design experiments with another 'real time' method of evaluation, action research. The paper concludes with some reflections on the scope and potential of design experiments.

Design experiments offer social scientists and public managers a new way of researching public policy innovations, which meet the highest standards of social science inquiry and supply timely information to decision-makers as the intervention develops. This method draws on the experimental tradition in social science, but is orientated to the micro-detail of the case rather than to the larger numbers usual in experimental evaluations, such as randomised control trials. The procedures of design experiments seek to overcome some of the barriers policy-orientated researchers face when evaluating public initiatives, notably the difficulty of feeding back the findings to core decision-makers and to front-line bureaucrats within a realistic timescale. Many standard methods face the trade-off between the speed of response and scientific validity; but the real-time response in design experiments increases inferential power rather than reduces it. With these advantages, design experiments open up many opportunities to researchers and policy-makers.

The paper aims to build on the advances achieved in original design experiments to allow for their wider application to a range of policy innovations. This route is needed because the current literature on design experiments is highly specialised and subject specific. They are based on evaluations of teaching practices to examine the efficiency of classroom-based learning techniques. This paper argues that the methods used in these studies can apply to a wide range of policy problems as they happen and constitute a general form of social science inquiry with wide application to the study of innovations in the public sector. Given the novelty of design experiments, the paper sets out the method and reports on how education researchers have applied them, and defends their validity in these studies. The paper then proposes some adaptations to the techniques used these education studies, with the

purpose of fostering their application to a range of policy fields and problem areas. Then the paper discusses the similarities and important differences with another 'real time' method for evaluating policy, namely action research. The argument is that, whilst design experiments share some common features with action research, their location in the experimental tradition offers advances in theory building, model testing and knowledge acquisition.

Design experiments in educational research

Researchers have developed design experiments to improve teaching practices in the US education system. They aim to be 'a form of interventionist research that creates and evaluates novel conditions for learning' (Schwartz et al 2005: 2). They mainly derive from Brown (1992) and Collins's (1992) seminal papers, which claim that researchers could depart from the classic features of the experimental method without breaching the high standards of social science enquiry. These early publications were followed by extensive theoretical commentary (e.g. Cobb et al 2002, The Design-Based Research Collective 2002, Edelson 2002, Bell 2004, Collins et al 2004), special issues of journals (*Education Researcher* 32(1), *Journal of the Learning Sciences*, 13(1), and *Educational Psychologist*, 39(4)) and several empirical studies in the learning sciences research community, some of which this paper sets out.

Design experiments need to be carried out at the outset of an intervention, making them high risk – they can easily go wrong – but they may be 'high gain' as well. The focus is on understanding and applying innovative techniques in teaching and learning. They aim for the 'engineering' of the learning experience (Cobb et al 2003). This use of this tag is deliberate because proponents of design experiments see

their origins in applied science, relying on a precise method that involves tinkering with the design of the intervention and learning from mistakes so the final product is finished to the highest standard. Just as with product design, the practitioners and researchers make minute adjustments to the specification, which are tracked in detail, and maybe further adjusted over time. As Collins et al (2004: 8) write, 'the design is constantly revised based on experience, until all the bugs are worked out.' The belief in the validity of the design experiment derives from a sceptical view of the efficiency of expert programmes of learning, such as those based on software developed in laboratories, which often fail when tried in the classroom; in contrast, those implemented in an iterative and incremental way can succeed (Hsi 1998). With the resulting best practice, the innovation can be introduced into other classroom settings. At the same time, it is highly likely that some of the aims of the instigators will shift as the design is implemented and adjusted (Brown and Campione 1996).

Design experiments are not atheoretical exercises. Although they do not use the language of recent policy evaluation and refer to a the 'theory of change' for assessing policy (cf. Pawson and Tilley 1997, Mackenzie and Blamey 2005), in practice they regard themselves as testing theories about how change will is taking place. Design experiments work with one or several theories, and develop a practical understanding about what could work, which can derive from social science theory and/or the intentions of those carrying out the intervention. Brown (1992: 143) writes, when contrasting design experiments to laboratory- based research, 'Even though the research setting has changed dramatically, my goal remains the same: to work toward a theoretical model of learning and instruction rooted in a firm empirical base … I attempt to engineer interventions that not only work by recognizable standards but are

also based on theoretical descriptions that delineate why they work, and thus render them reliable and repeatable'. The origins of many design experiment researchers in social psychology means that hypotheses from this discipline appear in these education design experiments, but researchers also apply hypotheses from their experience and the previous research of the design experimenter. Hypotheses also appear from other academic disciplines.

Design experiments explore and test hypotheses in a manner that is similar but not identical to classic experimental research. Instead of the once and for all hypothesis test, in the form of a statistical verification common in randomised control trials, which may confirm or reject the hypothesis, there is a quick turnover of research questions linked to the main hypothesis, followed by rapid redesigns of the experiment, what Cobb et al (2003: 10) call 'conjecture driven tests'. The theory is more practice driven than in classic social science – the idea is not to test general theories but to understand the practical limits and possibilities what the innovation is trying to do. Design writers (Collins 1992, Collins et al (2004) start with Herbert Simon's distinction in his book, *The Sciences of the Artificial*, between the type of theory building in the sciences of physics, biology, and that of the design or 'artificial' science, engineering and computer science, which are more recent and where theory is more linked to the tasks at hand. Thus the theory in design experiments is intermediate, standing somewhere in between the grand statements of educational theory and accounts of practical relationships on the ground (diSessa 1991).

Design researchers regard the classroom is an interactive space where the experiment takes shape in the form of a 'learning ecology' (Cobb et al 2003: 9). The classroom

setting is a system of interactions, whereby one small change affects all the other inputs and outputs. As Brown (1992: 143) writes, 'Just as it is impossible to change one aspect of the system without creating perturbation in others, so it is difficult to study any one aspect independently from the whole operating system. Thus we are responsible for simultaneous changes in the system, concerning the role of students and teachers, the type of curriculum, the place of technology, and so forth. These are seen as inputs into the working whole.' This integrated approach contrasts with laboratory-based studies of learning. Thus Collins (1999) sees design experiments as messy: there may be more than one dependent variable; it is not possible to introduce controls; there is 'flexible design revision' rather than fixed procedures; the subjects interact in their social world rather than being isolated as in a laboratory; the researcher develops a profile rather than directly tests hypotheses; and the participants participate in the design. Teaching styles and the composition of classes may change from day to day in the classroom, in a manner that would violate a classic randomised control trial design. As Brown writes, 'making this shift involves an increasing trade-off between experimental control and richness and reality. The classroom is not the natural habitat of many experimental psychologists, and our methods did not evolve to capture learning in situ.' (Brown, 1992: 152).

There is also a sophisticated understanding of scientific practice, which does not always involve experimentation, but much time preparing for an experiment, in the form of introducing and calibrating instruments, a process called instrumentation, with the claim that 95 per cent of scientists' time is spent on this way rather than on the experiments themselves. Not all design experiment theorists stress the difference with the laboratory-based approach, regarding them as complementary and

intersecting (Brown 1992, McCandiss et all 2003). As the Design-Based Research Collective writes (2003: 6), 'design-based research can generate plausible causal accounts because of its focus on linking processes to outcomes in particular settings, and can productively be linked with controlled laboratory experiments or randomised clinical trials … by assisting in the identification of relevant contextual factors, aiding in identification of mechanisms (not just relationships), and enriching our understanding of the nature of the intervention itself.' Methodologically, most design-based researchers are open, allowing for a range of methods and leave aside direct questions about epistemology (Bell 2004). Whether the design experiment can so easily escape the classic philosophical debate between schools of thought is a question this paper returns to.

The researcher's role resembles that of a participant. There is no reason why practitioners cannot be researchers, and many design experiments are practitioner-run projects; and where there are researchers, there is a very close collaboration. In fact, sometimes design experiments are known as design partnerships (http://www.soe.berkeley.edu/sandhtdocs/guide.html). In general, the person doing the design experiment participates in the innovation and in its evaluation. As Brown writes, 'As a design scientist in my field, I attempt to engineer innovative educational environments and simultaneously conduct experimental studies of those innovations' (1992: 141).

Finally, design experimenters think about their adoption in more normal environments without the direct participation of the researchers and policy advocates. This is the traditional problem of educational and other types of innovation: the contrast between

the early phase of the excitement of experimental innovation and often good results; then the disillusion of the application to more normal settings and the numbing effect of bureaucratic routines and policy-makers' competing priorities. Here design experiments have the edge because they roll out the innovation in the light of implementation issues and real world constraints. To ensure implementers reap the rewards, Brown (1992: 172) recommends that design experiments should have three stages: the alpha of development; the beta of tryouts at carefully chosen sites; and the gamma of wide spread adoption with minimal support – she draws these terms from software development.

Examples of design experiments in education

Design researchers usually focus on developing courses of learning, such as the BGuILE project of supporting materials for learning in biology (Reiser et al 2001) and the project based on science discussion in a Multimedia Forum Kiosk (Hsi 1998). And there are many projects based on Information Technology adoption, such as the GenScope software tool to help students explore ideas in genetics (http://genscope.concord.org/, Horwitz and Christie 1999, Hickey et al 1999, Hickey et al 2003). Here we describe two design experiments in detail. The first is Brown's Fostering Community of Learners (Brown and Campione (1994, 1996; Brown, 1992), the pioneering design experiment, which aimed to promote a 'development corridor', extending a learning community across the divisions of grades and classes, focused on particular topics like ecology. There were three cycles of activities in their design, beginning with students breaking into sets of research groups, but with cross-talk across them. The cycle ended with a concluding session. This design arose from the experiment, whereby participants designed the learning environment and researched it

at the same time, starting from an adaptation of reciprocal teaching, with the idea of the community of learners appearing after the first stage of the design experiment as the researchers observed students helping each other with the learning, using their diverse expertise. During the experiment they found that certain things went wrong, such as students learning the wrong sorts of biological facts, and they had to adjust the content of the initial materials and design sets of activities to produce particular kinds of learning outcomes. A second innovation the researchers introduced was an interactive learning outside the classroom through the World Wide Web. All these new problems and innovations led the researchers to modify the design. Brown and Campione gathered a vast amount of information during the research: the standard measures of reading, writing, content knowledge, and computer competence, all of which improved; and then information that was not readily measurable: transcripts of children's planning, revising and teaching sessions; observations of teachers' coaching and responsive teaching and direct instruction; records of student portfolios, including individual and group long-term projects; and scoring of electronic mail queries to peers, teachers, and collaborators in the university community. There were ethnographic observations of group discussions, planning sessions, students seeking help, peer tutoring and extensive video and audio taping of individuals, groups, and classroom settings. This data intensive character of design experiments makes them unique as well as the mix of methods.

Another example is Joseph's (2004) work on the *Passion Schools Project*, which used 'deep' learner interests to drive the learning experience. Students received curricula according to their interests and they learn actively with expert adults and with more and less advanced students. In the passion schools design experiment, students learn

to create films as part of a video crew. Phase one of the project took place in an urban fifth grade classroom in 1996 with 33 students. Joseph split the students into groups and assigned each one a role, which culminated in the production of the video. Joseph and several assistants observed the intervention, and interviewed participants prior and after the making of the videos. Then they redesigned the project to encourage students to focus on serious learning activities, creating a system of accreditation within the project, which was implemented in stage two in 1997, which involved smaller numbers and working outside school hours, and more focus on mentoring. Subsequently, the researchers became interested in theory, which influenced phase three in charter schools in 2000 and in 2002. The modifications were training for mentors and the showing of videos at different times, and then a subsequent phase, a new curriculum, the Multi-Media Design Studio introduced in 2002.

General principles for design experiments

So far design experiments occur only in schools and their findings appear in specialist and practically-orientated learning science journals. What scope is there for a more general application? It could be a simple case of transplanting the method to other contexts on the grounds that if it works in education, then it should work elsewhere. There two lines of resistance against using design experiments for researching policy innovations outside the classroom. The first is that they are too narrowly focused on classroom techniques, in particular on cognitive and learning processes, to be of wider interest. After all, it is the variable and practical environment of the classroom that led education researchers to reject more classical modes of evaluative methodologies. But is the rest of the world so different, with frequent changes to administration and

policy, and where street-level bureaucrats seek to implement innovations to affect outcomes around them? A lot of policy-making is in fact about the transmission of ideas and learning, such as ideas about drug-free living to people in treatment for substance abuse, whereas the more directive forms of public policy in the form of sanctions and incentives tend to be ineffective on their own. And the classroom too is a social world where restriction, incentives, learning and sanction all work together. While it is true that one of the main innovators of design experiments originated from cognitive psychology (Brown 1992), she moved away from it, and most design experiments have a broad social science interest in causation and outcomes. However, it remains fair to say that design experimenters who operate outside the classroom experience cannot have the same level of involvement and direction of their environment. It is just not possible to imagine researchers taking the role of a social worker or a police officer, in contrast to a teacher, for example. However, if the same amount of involvement is not possible in non-education environments, it is still possible to have more than in classical qualitative methods and replicate all the other routes of experimental enquiry. The second objection is that design experiments are not in fact scientific at all – they have abandoned some of the key elements of the experimental method, in particular control over the experiment, making it impossible to derive a causal inference. This criticism influenced Brown's work, and she was very careful to refute this sort of attack as many of the quotations from her work shows. But is it useful to set out the scientific principles of design experiments, partly to clarify the method and also to indicate in what contexts they are likely to be successful.

So what are the necessary features of a design experiment? Existing design experiments do not always achieve a full implementation of these guidelines. We identify nine features below. The first batch relate to the character and nature of a design experiment. The starting point has to be that theory must come first. It is not possible to dive into a design experiment without thinking through the causal linkages or with a hypothesis to test. That is not to say that, in the best traditions of science, that theory may not change as a result of empirical work, and in this case theory is crucially linked to practice in the way that engineering theory is different for theoretical physics. Second, it must involve a strong collaboration with policy-makers and those who deliver the policy on the ground. It may involve the researcher delivering the innovation directly, which allows the intensity of observation and direct control over the instruments of the intervention. This gives design experimenters a higher degree of control over their environment than many other social scientists have, closer to the experimental ideal. Third, the design experiment needs to be about a decisive intervention, which is highly innovatory. More standard methods would apply to more normal or modest shifts in direction for policy-makers that mainly replicate or extend existing practices, and where there is interest in the mean level of response: the evaluation of innovations needs attention to the particular case at hand, whether a failure or a success, and the efforts made to achieve success (Perrin 2002).

The next set of guidelines clarifies the need for intense and repeated data collection. So the fourth guide point is that, at the start of the project, throughout its course, and at its milestones and the end, there must be detailed recording of the intervention and of the views of those who are involved with delivering it. This requirement makes the

demands of the design experiment much greater than in the standard research project, because the recording of different sorts of detail is at the core of observing how different aspects to the intervention affect outcomes. In particular, it is important to record any discussion the researchers have with the policy-makers and front-line bureaucrats. This process is essential for tracing the causal processes and the precise link between outcomes and the calibration of the instruments. Brown (1992: 162) also thinks that the detail of recording tries to avoid the Bartlett effect, of the researcher only selecting those elements of transcripts, among the very large number of them, that prove her or his point, or selecting the 'golden moments' rather than the 'dross-like' ones. It may not be possible to equally evaluate all the information, but we should 'we continue to keep all notes, audio and videotapes, and transcripts on file so that our selection bias can be checked a later indefatigable sleuth, even though we believe this is unlikely to occur.' Fifth, design experiments proceed through iterations: there are cycles associated with the intervention, which naturally permit a review with the researchers, and then a redesign. Such a practice regularises the design-re-design process of a classic experiment. There is an acceptance that the experiment changes over time with more information about what aspects of the intervention work.

The next batch of guidelines focus on the challenge of designing appropriate measures. The sixth is that calibration and instrumentation are key features. The idea is that the minute aspects of the tools of government that are important in creating change, and that these need to be refined and recast in the light of experience. The seventh is that the precision of observation also requires a high degree of attention to construct validity, ensuring that what is being measured is accurately and recorded by the appropriate instruments. Even though the main character of the method is

qualitative, it usually needs the precision of quantitative measures to be able to show the results of a small change in design. Again the classic laboratory experiments are based on a very small number of cases, but where the external environment is manipulated. Of course, it is not possible to manipulate all aspects of the social world, but in principle it is possible to increase the leverage researchers and practitioners have when carrying out an intervention.

The last set of guidelines relate to what to expect in terms of output from a design experiment. Eighth, is that the intensity of observation requires a small number of settings where the experiment can be carried out, often in just one setting. It is not possible to imagine the design experiment happening across a large-n study design. Ninth, the experimenters tend not to evaluate 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.

These nine conditions are essential for the intervention to qualify as a design experiment. But there is one additional condition that enhances the quality of the research, which does not appear in the education research, and may be regarded as a modification or extension of the method. As with a quasi-experimental research design, it is important to have a control group or area, where there is a different intervention or more standard form of practice, which is observed in much the same way as before (though with less attention to the practitioners and policy-makers on the ground). Here it is possible to contrast the design experiment with what is happening in the absence of the experiment, though in a way that does not meet the rules of

statistical inference. It also helps avoid the problem of a Hawthorne effect with the design experiment, which can be a critical problem with the method, though some Hawthorne effects are inevitable and need to be taken into account when transferring the policy design. Brown (1992: 164-5) believes Hawthorne effects to be integral to the intervention – without interest and engagement on the part of the students, the intervention will not work: 'the Hawthorne effect is exactly what I am aiming for in my classrooms' (1992: 165). But if the control is also an intervention, but one where the techniques are very different, perhaps with another 'do nothing' control, it may be possible to separate out the Hawthorne element from other aspects of the intervention.

Overall, these conditions create a plausible form of social investigation that meets the highest standards of social science enquiry by being theory-driven and paying rigorous attention to measurement. What is different from standard social scientific practice is that the experiment is allowed to change over time and to react to its environment. Rather than seeing this as something to be avoided in the interests of controls, the variations under experimental conditions are what allow the researcher to trace the causal processes. That said, the design experiment should avoid massive non-relevant confounds to the experiment or else there may be so much variation to make it entirely impossible to carry out the research.

Much of this article takes the form of an argument. By summarising what is best in the education design experiment tradition, the idea is that it could be extended to other kinds of activities. However, there are the first signs of an experimental approach being developed in the UK. There is the start of a wave of design experiments being commissioned, with the pioneer being The Trees Estate Project, a collaboration

between Manchester University and Wigan Metropolitan Borough Council in a design experiment about a new initiative to combat substance misuse on one of its estates. The authority's new approach to drug treatment is based on intensive support of a client group and on building community relationships; it is not based on methadone prescribing, what they call 'the social approach'. People in the treatment programme have intensive support from a dedicated worker, have access to medical support, training, sports activities and holidays, with the idea of seeking to create meaning in their lives, but without drug use. Researchers at the university are involved with the design of the project, and also researching a 'control group' on a very similar estate where there is existing interventions. The method involves monthly meetings with policy-makers, a questionnaire based measurement of outcomes, and detailed monitoring of the intervention, including participation in the programmes and working alongside the case workers. The project started in September 2005 and concludes in 2007. Further projects may be commissioned by the North West Improvement Network.

How different are design experiments to action research approaches?

It is fairly obvious that design experiments are different from most sorts of evaluative methods, such as standard statistical tests of change over time and randomised control trials. It is also evident that design experiments are very special kind of qualitative research, and involve much more on the part of the researcher and practitioner than a case study or an analysis of stakeholder views or ethnographic methods. But it may be the case some forms of evaluation are similar to design experiments, particularly as closeness to policy-makers and the refinement of an intervention are common sense proposals. Action research is just such an iterative methodology that relies on cycles

of direct interaction with the decision-makers and on refining how policy-makers intervene. Invented by Kurt Lewin in the 1940s, it is associated with the experience of practice: 'The research needed for social practice can best be characterized as research for social management or social engineering. It is a type of action-research, a comparative research on the conditions and effects of various forms of social action, and research leading to social action' Lewin 1946, reproduced in Lewin 1948: 202-3). Lewin planned there would be a spiral of steps, 'each of which is composed of a circle of planning, action and fact-finding about the result of the action' (*ibid.*: 206), which again is similar to the design experiment method, with the idea of modifying the plan of action, which may be seen as similar to design-based research.

However, it would be wrong to conflate action research and design experiments just because they share some aspects. One clue to the difference is that doing action research is associated with a certain kind of stance toward politics and society, often critical of standard approaches to policy formulation and evaluation that, so it is claimed, replicate historic power relations. Action research implies action to remedy the social or bureaucratic problems. In the 1960s this association between doing research and having radical political values was highly explicit, and contributed to the decline of action research after that period. In spite of increasing rigour and care in application in later years, when the method again become popular, the location of action research in community work and participatory kinds of policy work tends to maintain this value commitment. Action research is thus not just about an iterative approach. Design experiments, on the other hand, owe nothing to a set of values and emerge solely from practical concerns, and can concern any kind of innovation, with the normative argument being carried out largely elsewhere. Action researchers by

contrast are inclined to stress the views of those who are subjected to a policy innovation, such as the participants in a housing scheme for example, tending to stress that the method is democratic and liberating (Stringer 1999). Such a stance tends to overlap with action research's epistemology, which is often developed in opposition to dominant positivist methodologies (Kock et al 2005). The claim is that research is only possible if the researcher is part of what is being researched. There is a stress on critical reflection, rather than mainly observation. Again design experiments, while doing something superficially similar, is in fact rather different: the closeness to the researcher is not about empathy and getting into the minds of the researched, but is about replicating the closeness of a scientific or engineering design. After all, the detail with which the engineer in involved with the design is about getting the micro-detail right, which is essential for the design, and it is the same with design experiments. These contrasts are not necessary, and in fact it is possible to carry out action research in the same manner as design experiments (see Koch et al 2005), but these modifications would take action research away from its core concerns. In this sense, to rephrase an aphorism: design experiments and action research are 'alike in unimportant respects'.

To make the point, it is worth summarising the similarities and differences between action research and design experiments in a table. Here the basic elements of research methods are along the stem, with the two methods compared side by side. By examining these contrast it shows the sharing of features, such as the collaboration and iteration, but the differences stress design experiment's location in the scientific tradition, the precision of measurement and an epistemology within the mainstream of scientific practice rather than being critical of standard empiricist accounts.

Table: common features to action research and design experiments

	Action Research	Design Experiment
Theory based	Not necessary	Necessary
Research collaboration	Necessary	Necessary
Iterative approach	Necessary	Necessary
Precision of measurement	Not necessary	Necessary
Value commitments	Likely	Outside the research
Reflection on practice	Necessary	Subsidiary
Epistemology	Anti-positivist	Positivist or realist

Conclusions

Design experiments offer researchers a new form of investigating interventions. They are a way of introducing evidence into the policy making process that recognises and copes with some of the inherent contingencies and qualities of the world of policy-making. Yet in the commitment to theory, the rigour and care of data collection and the focus on producing clear and applicable findings the highest commitment to scientific research principles are maintained. We argue that design experiments can help in the shaping of a policy by producing warrantable knowledge that in turn can match a set of scientific principles on which many researchers would agree (National Research Council, 2002) As social scientists value the production of evidence, the linking of evidence to theory, scepticism about claims to knowledge, a keenness to

rule out alternative explanations and a commitment to constructive criticism as key ingredients in developing knowledge. Design experiments deliver on all those elements.

They are an advance on existing methods, such as action research and case studies because they combine aspects of research that are usually separated – making qualitative research as rigorous as possible by replicating the experimental method. By making collaboration with researchers a value in observing the experiment and in feeding back alteration to the design there are observable benefits for the policy-maker. By concentrating on what works to the exclusion of everything else, the value to the practitioner is only matched by the purity of the scientific endeavour.

Design experiments are not suited to all sorts of evaluation. They are not appropriate for routine work or large scale programmes, where the precision of design experiments cannot be applied. They are best when examining innovations rather than roll out or re-specification of programmes. Design experiments favour the small scale innovation, in a relatively controlled environment, where the dialogue can take place with a small range of policy-makers and workers, all of whom have signed up to the new way of doing business and to intense researcher-practitioner interactions. It suits a small scale, spatially concentrated locations, but not quite introduced with the fanfare and public scrutiny of a pilot area, where it is assumed there is a wider programme ready to be rolled out. But many of these limitations of the design experiment method are in fact its advantages. Only certain kinds of policy-makers and practitioners can come together with the researchers in some contexts, which avoids the problem of overload and too high expectations. By being more specialised,

and targeted to the innovators, it encourages a high degree of focus and commitment on those who participate. It also biases the role of the policy-maker to try our more innovative ways of intervening. The payoff is that the causal mechanisms can be understood much more clearly in relationship to the instruments the policy-makers have to hand, and the possibility that other organisations and deliver bodies wishing to implement the design will not need to start from scratch when implementing it, and will not need the experimental approach set out here.

Currently, there are few design experiments outside educational practice, and even these educational projects have not had the attention they deserve. But the claim is that the potential is there for both the social science and practitioner communities to reap. Our prediction is that design experiments will emerge from their productive location in education research to make a useful contribution to the evaluation field.

References

Bell, P. L. (2004). 'On the theoretical breadth of design-based research in education'. *Educational Psychologist*, 39(4), 243-253.

Brown, A. (1992). 'Design experiments: theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141-178.

Brown, A. L., & Campione, J. (1994). 'Guided discovery in a community of learners'. In K. McGilley (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229-270). Cambridge, MA: MIT Press/Bradford Books

Brown, A., and Campione, J. (1996). 'Psychological theory and the design of innovative learning environments: On procedures, principles, and systems'. In L. Schauble & R. Glaser (Eds.)*Innovations in learning: New environments for education* (pp. 289-325). Mahwah NJ: Lawrence Erlbaum Associates.

Cobb, P. Confrey, J., DiSessa, A. Lehrer, R, and Schauble, L. (2003) 'Design experiments in educational research' in A.Kelly (ed) *Educational Researcher* 32, 9-13.

Brown, A., Ellery, S., & Campione, J. (1998) Creating zones of proximal development electronically. In J. G. Greeno and S. Goldman (Eds.) *Thinking practices in mathematics and science learning* (pp. 341-367). Mahwah NJ: Lawrence Erlbaum Associates.

Collins, A. (1992), 'Towards a design science of education', In Scanlon, E. and O'Shea, T.(eds.) *New Directions in Educational Technology*, Springer Verlag, Berlin.

Collins, A. (1999) The changing infrastructure of education research. In E. Lagemann & L. Shulman (Eds.) *Issues in education research* (pp. 289-298). San Francisco: Jossey-Bass.

Collins. A., Joseph, D. and Bielaczyc, K. (2004), 'Design research: theoretical and methodological issues', *Journal of the Learning Sciences*, 13(1): 15-42.

diSessa (1991) 'Local Sciences: Viewing the Design of Human-Computer Systems as Cognitive Science', in *Designing Interaction: Psychology at the Human-Computer Interface*. NY: Cambridge University Press, 162-202.

The Design-Based Research Collective (2002), 'Design-based research: an emerging paradigm for educational inquiry', *Educational Researcher*, 32(1): 5-8.

Edelson, D. C. (2002). 'Design research: What we learn when we engage in design' *Journal of the Learning Sciences*, 11(1), 105-121.

Green, D. and Gerber, A. (2002), 'Reclaiming the experimental tradition in political science', in H. Milner and I. Katznelson (eds.), *State of the Disclipline*, Vol 111 (New York: Norton.

Horwitz, P., & Christie, M. (1999). 'Hypermodels: Embedding curriculum and assessment in computer-based manipulatives'. *Journal of Education*, 181(2), 1-23.

Hickey, D.T., Kindfield, A.C.H., Horwitz, P., and Christie, M. (1999). 'Advancing educational theory by enhancing practice in a technology supported genetics learning environment.' *Journal of Education*, 181(2), 24-55.

Hickey, D. T., Kindfield, A. C. H., Horwitz, P., & Christie, M. A. (2003). 'Integrating curriculum, instruction, assessment, and evaluation in a technology-supported genetics environment'. *American Educational Research Journal*, 40 (2) 495-538.

Hsi, S. (1998), 'The multimedia forum kiosk: using design experiments to understand electronic scientific discussions', unpublished paper.

http://kie.berkeley.edu/events/hsi-aera98.pdf

Joseph, D. (2004), ‘The practice of design-based research: uncovering the interplay between design, research, and the real-world context’, *Educational Psychologist*, 39(4): 235-242

Koch, N., McQueen, R. and Scott, J. (2005), ‘Can action research be made more rigorous in a positivist sense? The contribution of an iterative approach’, http:/www.scu.edu.au/schools/gcm/ar/arrow/kms.html

McCandiss, B., Kalchman, M. and Bryant, P. (2003), ‘Design experiments and laboratory approaches to learning: steps toward collaborative exchange’, *Educational Researcher*, 32(1). 14-16.

Mackenzie, M. and Blamey, A. (2005), ‘The practice and the theory: lessons from the application of a theories of change approach’, *Evaluation* 2005 11: 151-168.

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

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

Perrin, B. (2002), **‘**How to - and how not to - evaluate innovation’ *Evaluation* 2002 8: 13-28.

Reiser, B., Tabak, I., Sandoval, W., Smith, B Steinmuller, F., & Leone, A. J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 263-305). Mahwah, NJ

Schwartz, D., Chang. J. and Lee, M. (2005), ‘Instrumentation and innovation in design experiments: taking the turn towards efficiency’, unpublished paper

Stringer, E. T. (1999) *Action Research: A handbook for practitioners 2e*, Newbury Park, Ca.: Sage.