

Newcastle University e-prints

Date deposited: 11th August 2010

Version of file: Author, final

Peer Review Status: Peer reviewed

Citation for published item:

Dyson A, Todd L. [Dealing with Complexity: Theory of Change Evaluation and the Full Service Extended Schools Initiative](#). *International Journal of Research and Method in Education* 2010,**33** 2 119-134.

Further information on publisher website:

<http://www.tandf.co.uk/>

Publishers copyright statement:

This paper was originally published by Taylor and Francis, 2010 and can be seen from the URL below (with permissions)

<http://dx.doi.org/10.1080/1743727X.2010.484606>

Always use the definitive version when citing.

Use Policy:

The full-text may be used and/or reproduced and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not for profit purposes provided that:

- A full bibliographic reference is made to the original source
- A link is made to the metadata record in Newcastle E-prints
- The full text is not changed in any way.

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Robinson Library, University of Newcastle upon Tyne, Newcastle upon Tyne.

NE1 7RU.

Tel. 0191 222 6000

**Dealing with complexity: theory of change evaluation
and the full service extended schools initiative**

Alan Dyson, School of Education, University of Manchester

**Liz Todd, School of Education, Communication and Language Sciences,
Newcastle University**

Address for correspondence

Professor Liz Todd
School of Education, Communication and Language Sciences,
King George VI Building
Newcastle University
Queen Victoria Road
Newcastle upon Tyne, NE1 7RU

tel 0191 222 6572 / 222 5155
liz.Todd@ncl.ac.uk

Dealing with complexity: theory of change evaluation and the full service extended schools initiative

Summary

The full service extended schools initiative in England sought to build the capacity of schools to deploy multiple interventions across a range of child, family, community contexts in order to improve the lives of children and families, and to enhance the sustainability of the areas where they live. As such, it was part of wider international developments to broaden the role of schools and integrate their work with that of other child and family services. The complexity and indeterminacy of such initiatives presents particular challenges to evaluators, and mean that simple input-output approaches to evaluation are inappropriate. This paper reports the use of a theory of change approach to the full service extended schools evaluation. It argues that an approach of this kind is better able to deal with complexity. However, in practice, this approach did not simply replicate standard outcomes evaluation processes in a more context-sensitive way. Rather, it involved fundamental shifts in what counted as valued outcomes, in a focus on potential rather than effectiveness, and in the roles of actors and evaluators. The paper concludes that, as the role of schools widens, educational evaluators will also have to rethink their roles.

Dealing with complexity: theory of change evaluation and the full service extended schools initiative

Background

Since coming to power in 1997, New Labour governments in England have pursued what former Prime Minister, Tony Blair, characterised as “an unprecedented crusade to raise standards” (Blair 1999). In line with developments in the USA and many other countries, these standards-based reforms have focused on raising school students’ measured attainments through a combination of high-stakes testing and accountability, central prescription of the standards to be achieved (and, indeed, of curriculum content and teaching methods), and successive government initiatives targeted at low-performing schools and student groups (Ball 2008; Wyse, McCreery, and Torrance 2008; Chapman and Gunter 2009).

Integral to these reforms has been the use of data from national assessments to monitor the performance of the school system as a whole. The National Pupil Database (NPD) brings together attainment records and demographic data for all pupils in the school system, making it possible to track the performance and progress of individuals, groups and institutions, and to explore the relationships between outcomes and background variables (see Schools Analysis and Research Division 2009 for a recent example). The NPD also simplifies the process of evaluating the outcomes of even quite complex interventions. Given the availability of such extensive outcomes data, it becomes a *relatively* straightforward matter to explore

outcomes pre- and post-intervention, or in intervention and comparator sites (see, for instance, the recent evaluations of the Excellence in Cities initiative, NFER 2007), or, more contentiously, of the academies programme (National Audit Office 2007; Gorard 2005; Pricewaterhouse Coopers LLP 2007).

However, not all New Labour education policies have shared this straightforward – some might say, narrow – focus on ‘standards’. Other strands have seen educational attainment as just one amongst a wide range of desirable childhood outcomes. They have therefore understood these outcomes as being shaped by the child’s wider social environment, and supported by the full range of child and family services, rather than simply by the school. An overarching policy framework for this work has been provided by the Every Child Matters initiative (DfES 2003a) which seeks, amongst other things, to create integrated child and family services at local level to pursue five shared outcomes – be healthy, stay safe, enjoy and achieve, make a positive contribution, and achieve economic well-being.

These wider policy strands have some important consequences for how the role of schools and the outcomes of schooling are understood. Instead of being more-or-less simple machines for raising measured attainment, schools become part of a complex network of services aimed at producing a wide range of interlinked outcomes for children, their families, and their communities. This has obvious implications for accountability processes. When proposals emerged recently for extending the range of outcomes for which schools were held to account, one teacher’s leader complained bitterly that:

...schools would be accountable to Ofsted [the schools' inspectorate] for the number of pregnancies - and for eighteen other things apart from learning achievements, including taking illegal drugs, being safe from crime outside school, engaging in decision-making in the community, and being law-abiding.

(Johnson 2008)

Much the same could be said of the evaluation of school-based interventions. If schools have a wider role than simply the raising of attainment, then evaluation has to take account of this wider role. To some extent, this is already happening. Under the aegis of Every Child Matters, school level interventions targeting outcomes other than attainment have begun to emerge, and evaluators have accordingly had to search for outcomes data well beyond that held in the National Pupil Database (see, for instance, (Barnard et al. 2009; Humphrey et al. 2008). However, simply extending the range of outcomes measures may not be enough. Experience beyond the school sector – most notably in the Sure Start initiative (DfEE 1999b) – suggests that multi-strand interventions in complex contexts may require at the very least a corresponding increase in the complexity of evaluation designs (Belsky, Barnes, and Melhuish 2007; National Evaluation of Sure Start no date, Toynbee 2005). As yet, there is only limited experience in England – and, arguably, internationally – in managing this increased complexity in respect of interventions in and around schools.

The remainder of this paper, therefore, is concerned with the authors' experiences in leading the evaluation of a complex intervention of this kind. The national 'full service extended schools' initiative was an attempt to build the capacity of schools to deploy multiple interventions across a range of child, family, community contexts in

order to improve the lives of children and families, and to enhance the sustainability of the areas where they live. In what follows, we describe the characteristics of the initiative and the challenges this posed to the evaluation. We offer an account of our use of a methodology – theory of change approach – which met at least some of these challenges, and note the surprising directions we were taken by the methodology. Above all, we consider what the implications of our experience might be for school-focused evaluation at a time when the role of schools is changing.

Challenges in evaluating full service extended schools

In 2003, the Department for Education and Skills (DfES) - the government department then responsible for education in England - launched its full service extended schools (FSES) initiative. This was aimed at supporting the development in every local authority area of one or more schools providing a comprehensive range of services on a single site to their students, to students' families, and to local communities. These services were expected to include childcare, health and social care, lifelong learning, family learning, parenting support, study support, sports and arts provision, and information and communications technologies provision (DfESa, b: 3-4). The authors led a team commissioned by DfES to undertake an evaluation of this initiative, with field research between 2004 and 2006. Interim reports were produced for the first two years of the evaluation, with a final report published in 2007 (Cummings et al 2005, 2006, 2007) While the FSES initiative was still running, the Government announced a national 'roll out' of some (though by no means all) of its features so that every state-funded school in the country would be expected by 2010 to provide access to 'extended' services (DfES 2005).

These developments can be seen as part of an international movement towards reformulating the role of schools as bases and coordinating points for child, family and community services (Dyson in press). This movement has in turn generated an evaluative literature which promises to identify the impacts of this reformulated role (see, for instance, Blank, Melaville, and Shah 2003; Clark and Grimaldi 2005; Henderson and Mapp 2002; Sammons et al. 2003; Szirom et al. 2001). However, the promise of this literature is by no means always fulfilled. As a US-based review comments, somewhat ruefully:

It seems intuitively obvious that creating a context that interweaves home, school, and community, and that makes students valued and contributing members should have a powerful effect on student learning. But attempts to connect community collaborations and student test scores have been few and contradictory.

(Keyes and Gregg 2001: 40)

Similarly, a UK-based review undertaken to inform the development of extended schools, concludes that there has been “little systematic, rigorous evaluation of the concept and its implementation” (Wilkin, White, and Kinder 2003: 5). When we look at individual evaluations, we can see some of the challenges evaluators face. An evaluation of a Chicago full service schools initiative (FSSI), for instance, found some evidence of improved student achievement outcomes. However, these varied between participating schools and could not reliably be attributed to the initiative because:

Many efforts were underway in each school during the FSSI period to improve student outcomes, and no simple causal links can be drawn between FSSI and improvement at the three schools.

(Whalen 2002: 2)

In much the same way, the evaluation of the Children's Aid Society Community Schools in New York, amongst much evidence for process changes, found some evidence of improvements across a range of outcomes for students but these seem not to have been consistent or to be unequivocally attributable to the full service initiative.

The evaluators conclude that:

...community schools are complex systems making fundamental institutional changes, and the means that events occur in many ways and on many levels.

It is, therefore:

...necessary to look beyond standardized test scores to understand the impact of community schools.

(Clark and Grimaldi 2005: 173)

Amongst the many challenges facing evaluators in this field the following seem to be particularly significant:

1. Weak specification of the intervention

There is no single model of 'full service schooling' that is adopted by all schools and which, therefore, can reliably be evaluated (Ball 1998; Dryfoos and Maguire 2002; Wilkin et al. 2003). On the contrary, schools either develop their own practices idiosyncratically, or adapt programmes that are loosely specified by their sponsoring governmental or non-governmental organisation. This was certainly the case in the FSES initiative, where DfES specified the types of activity in which schools should engage, but not the precise form of these activities, or the weighting to be given to each (DfES 2003a, b). As a result, schools tended to develop packages of provision that differed from each other in important ways (Cummings et al. 2005).

2. Indeterminacy of outcomes

Although the sponsors and advocates of full service schools tend to promise much, the intended outcomes of these initiatives frequently take the form of lists of loosely-connected aspirations. DfES, for instance, suggested that FSESs might deliver: better access to local services; reductions in health inequalities; reductions in adult unemployment; improved staff recruitment and retention; enhanced partnership with ‘the community’; better supervision of children outside school hours; improved school security; greater parental involvement; improved local career development opportunities; improved student behaviour, attendance, participation, aspirations and attainment; increased student motivation and self-esteem; and better access to specialist support for students (DfES 2003b: 3). Schools often added more ambitious aspirations in terms of changing community cultures or bringing about a re-engagement with learning in the area (Cummings et al. 2005). Even if all of these outcomes might reasonably be expected to materialise, tracking them, tracing the interactions between them, and finding robust indicators of their presence would constitute a formidable challenge, particularly given the time-limited nature of the evaluation.

3. The complexity of the context

Full service initiatives tend to be located in schools struggling with the significant problems arising from the turbulent and disadvantaged communities they serve (Dryfoos 1994) and the cumulative effects of successive social and economic policy initiatives. In the contemporary English context, this means that they are typically the site of multiple other initiatives in and beyond the school, all of which might be

expected to have some effect on child, family and community outcomes.

Disaggregating the impacts of these initiatives is a major challenge. In the FSES initiative, for instance, most of the schools were also participating in a ‘Behaviour Improvement Programme’ (<http://www.dfes.gov.uk/behaviourimprovement/>) that, amongst other things, brought together multi-agency teams to work with children in schools, and thus replicated an important strand of full service provision. In practice, FSES leaders found it hard to differentiate between the full service and behaviour improvement aspects of their provision.

4. The lack of controls and comparators

Given the weak specification of full service school models and the complexity of the contexts in which they develop, it was often difficult to differentiate clearly between schools which did or did not offer full service provision, or to identify a historical point at which schools ‘become’ full service schools. FSESs were distinctive in reception of DfES funding, but they also developed in a context where many schools already displayed elements of full service provision (Ball 1998; DfEE 1999a; Wilkin et al. 2003) and where there had been a succession of previous extended school initiatives. To complicate matters, FSESs tended to serve particularly disadvantaged populations, which made comparisons with other schools never more than approximate (Cummings et al 2007).

Theory of change evaluation

Although these challenges to evaluation are to some extent attributable to avoidable weaknesses in intervention design and implementation, that is by no means the whole story. Arguably, attempts by schools to marshal multiple interventions in the

interactions between children, families and communities must pay attention to the complexities of particular situations, and must involve a degree of uncertainty as to outcomes. In other words, indeterminacy is not – or not simply – a function of poor design and implementation, but of the more complex roles that schools are playing. In any event, simple input-output models of evaluation, assuming that the financial, human and intellectual resources of an initiative will lead straightforwardly to readily-identifiable student outcomes, are unlikely to shed much light on initiatives of this kind. Indeed, when the Scottish version of full service schools – so-called ‘New Community Schools’ – were evaluated in this way shortly before the FSES initiative was launched, the unsurprising result was that few outcomes of any significance could be identified (Sammons et al. 2003). The complexity of schools working in the arena of a range of government initiatives mitigates against the application of more traditional, linear models of evaluation (Morrison 2002 and Brooke-Smith 2003).

Our solution to this problem was two-fold. First, like the Sure Start evaluation (Start no date), we opted for a multi-strand design working at both local and whole-initiative level. So, we undertook an analysis of pupil-level and school-level attainment data in FSESs and other schools, carried out a cost benefit analysis of FSES provision, surveyed head teachers, teachers, staff, parents and pupils in FSESs, and undertook more detailed case studies of implementation in a sample of FSESs, comparing them with a sample of schools that were similar in relevant respects but were not part of the FSES initiative (see Cummings et al 2007) for a more detailed account of evaluation design).

Second, we adopted a theory of change approach as the overall form of the evaluation. Theory of change is one of a family of theory-based evaluation approaches (Stame 2004; Weiss 1995) that start from the assumption that purposeful activity implies a ‘theory of action’ (Argyris and Schon 1978, 1996) in the form of a set of assumptions about the nature of the context in which action is set, the outcomes that are desirable in that context, and the way in which the action taken will lead to those outcomes. As an evaluation strategy, the theory of change approach involves, ‘a systematic and cumulative study of the links between activities, outcomes and context of the initiative’ (Connell and Kubisch 1998: 16). Typically, therefore, evaluators work with actors to explicate the latter’s underpinning theory or theories. Once this theory is articulated, evaluators can:

- examine the inherent coherence of the theory – that is, the extent to which a consensual theory can be articulated by actors, which offers a convincing account of how actions in context will produce intermediate changes that will generate intended outcomes;
- collect data on the intermediate changes produced by action, and assess the extent to which these changes occur as predicted by the theory; and
- collect data on any longer term outcomes that are generated in the evaluation period.

Two features of a theory of change approaches are worth emphasizing. First, they are not as dependent as many other forms of evaluation on being able to identify the end-point outcomes at which interventions aim. Although they certainly note such outcomes where they are available and can be combined with more classical approaches which focus exclusively on end-points (Weitzman 2002), they rely on

predicting what outcomes might emerge as much as on identifying outcomes that are already apparent. Outcomes in theory of change evaluations are conceptualised as materialising at the end of a chain of intermediate changes which the evaluation process seeks to track. At each point, therefore, evaluators are in a position to predict the outcomes that are likely to materialise in future and to assess the extent to which these predictions match those by the initiative's leaders. This means that the evaluation can give early feedback to those leaders and can provide an assessment of the initiative's likely outcomes even if, as is often the case, the evaluation has to conclude before they are fully apparent.

Second, the outcomes to be studied are not specified separately from the articulation of the theory of change as a whole. Rather than being imported into the evaluation by the evaluators or some external sponsors, outcomes are articulated by the actors in the initiative, albeit in negotiation with the evaluators (Weiss 1995). This of course raises issues about whose version of intended outcomes is to carry most weight, but at least this issue is surfaced and, where theories conflict, there is the option of studying different sets of outcomes in order to test different theories.

Theory of change approaches are labour intensive in working towards the initial articulation of their theory and in tracking what might well be lengthy and complex sequences of changes. They probably offer little, therefore, in the evaluation of tightly bounded initiatives with clearly identifiable outcomes. However, they come into their own where initiatives are complex, indeterminate, and set in complex contexts (Connell and Kubisch 1998). This is particularly the case, for instance, in situations like that facing us in the evaluation of FSESs, where initiatives are multi-strand and/or

variable in form, where intended outcomes are multiple, or poorly specified, or contested, and where the context of implementation contains many uncontrolled variables that might impact on the initiative.

As a further benefit, the dialogical nature of theory of change evaluations means that actors can receive continual feedback from evaluators on the coherence of their theories and on the unfolding of intermediate changes, long before end-point outcomes begin to emerge (Connell and Klem 2000). This means that changes to the initiative can be made in good time and the articulation of the theory adjusted accordingly. Indeed, it is a short step from this to using theory of change as a planning tool, where it begins to have much in common with logic modelling (see, for instance, W.K. Kellogg Foundation 2004).

Not surprisingly, then, theory of change evaluation is now widely used in complex contexts (Coote, Allen, and Woodhead 2004). In the UK, for instance, it has been used *inter alia* in major national evaluations of Local Strategic Partnerships (Geddes 2006), the Children's Fund (Edwards et al. 2006), and Health Action Zones (Sullivan, Barnes, and Matka 2002). Its use in educational contexts in this country would appear to be less common. This may simply be a function of the origin of recent forms of theory of change evaluation in complex community initiatives in the US. However, it may also be the case that, in England in particular, the complexity of many educational initiatives has been somewhat hidden by the ready availability of apparently robust outcome measures – in the forms of student attainment and school performance data – and the determination of policy makers that initiatives should, above all else, have an impact on these measures.

Theory of change in action

As Connell and Kubisch (1998) argue, the use of theories of change is an approach to evaluation rather than an evaluation method. Not only does it require the flexible adoption of methods from other forms of evaluation (including classical outcomes evaluations), but its broad principles can be and are applied in different ways to suit different evaluation contexts. In this section, we set out how we used this approach in the evaluation of the FSES initiative.

The national FSES initiative ran during the school years 2003-4 to 2005-6. Our evaluation worked with 17 FSESs between early 2004, shortly after the national initiative began, until late 2006, shortly after it had ended. Our work consisted of a series of phases of activity, intended first to articulate the school leaders' theories of change and then to shape the collection of data in relation to those theories. We relied throughout on recurrent interviewing, with head teachers and key school personnel involved in developing the full service approach. These interviews were supplemented with documentary analysis and with interviews with non-school personnel (such as local authority officers, community leaders and representatives of other community agencies) who might have different views of the needs of local people and the potential contribution of the school. As the theory of change became clearer in each school, the scope of the fieldwork widened to include the collection of evidence on the impacts that the school's full service provision was beginning to have on children, families and communities. Full details of methodology and sampling are provided in our end of year one report (Cummings et al. 2005).

We encountered some particular challenges in attempting to use theory of change methodology. On entering the field, we were prepared to encounter theories of change that took the form of complex articulations of the assumptions underpinning complex initiatives. We were also prepared for the business of articulating underlying theories to be what Mackenzie and Blamey describe as, “a slow and difficult process for the evaluators and many of the implementers” (Mackenzie and Blamey 2005: 163).

However, the reality was more difficult than even this might suggest.

Theory of change methodology rests on the apparently reasonable assumption that::

social programmes are based on explicit or implicit theories about how and why the programme will work.

(Geddes 2006: 6)

Our early work with school leaders suggested that this was only partly true in the case of the FSES initiative. Not only was that initiative loosely specified at government level, the interpretations of it by different schools themselves struggled for coherence. Schools’ responses tended to be a mix of purpose-designed actions, established actions rebadged as part of the full service approach, and actions developed opportunistically, often on a one-off basis. Actions emerged and disappeared in rapid succession, and it was not unknown for schools to ‘start again’ with a new approach. Frequently, disparate actions were developed by different staff members without any strong central control. Often, school leaders found it difficult to characterise the issues their FSES approach was intended to address or the outcomes it was intended to produce other than as a loosely-connected problems and possibilities, and they found it particularly difficult to say where the boundary lay between their FSES approach and other actions they were taking.

We had anticipated that our work with school leaders would be to make explicit a theory that was fully-formed but partially implicit. Accordingly, we had envisaged ourselves following Weiss' (1995) advice to map out theories 'in fine detail', tracing the complex links between actions and intended outcomes. Instead, we found ourselves spending a good deal of time filtering 'noise' out of the system, trying to identify not only what the assumptions were underpinning the school's FSES approach, but also which of the multiplicity of actions were and were not underpinned by those assumptions.

Insert figure 1 about here

Figure 1 presents a diagrammatic example of the outcome of this process in one school. In this case, school leaders had given us a typically long list of starting problems – high unemployment, poor housing, high levels of drug-related crime, low educational attainment – and an equally long list of intended outcomes – community learning, raised pupil attainments, positive school impacts on home, and so on. Whilst school leaders were happy to present and, indeed, extend these lists, they found it much more difficult to say how the starting problems were related to each other, which were causal and which were symptomatic, and which were priorities if the situation were to be changed. Likewise, they struggled to articulate the relationship between outcomes, or between outcomes and starting problems. Here, as in many other FSESs, what seemed like serious starting problems (in this case, high unemployment and drug crime were examples) seemed to have no corresponding outcome, whilst leaders constantly conflated 'outputs' in the sense of changes in their

own services and practices and ‘outcomes’ in the sense of benefits for service users (W.K. Kellogg Foundation 2004).

Between these problematic characterisations of the starting context and intended outcomes leaders described a set of actions the school was taking or intended to take – ranging from the establishment of an on-site health clinic to fast-track curriculum arrangements, adult education classes, the employment of parents as classroom assistants, and the provision of crèche facilities. Whilst the likely immediate benefits of these actions were not difficult to see, leaders found it difficult to look beyond these immediate benefits so as to link actions to starting issues on the one hand and longer term outcomes on the other. The reality was that these actions had not been planned strategically. Rather, an energetic head teacher and FSES coordinator had busied themselves making links with other agencies, talking to community groups, working with parents, and pursuing funding wherever they could find it, and offering the use of the many spare spaces in the school’s somewhat aged accommodation provided. What resulted was an eclectic mix of activities, only loosely tied to any attempt at tackling issues in the starting context or any coherent effort to generate specified long term outcomes.

Our work in this school, therefore, did not feel much like the ‘surfacing’ of a theory of change which, although implicit, was lying ready-formed for us to discover. Rather, we were in the business of helping leaders articulate some broad assumptions which guided their work in general terms but which were not tightly articulated and which left ample room for changes of direction, opportunistic action, and action whose purpose was far from clear. These assumptions typically represented not a detailed

explication of the theory of change, but a simplification of the mass of unconnected detail presented in our initial encounters with leaders into the much broader statements set out in figure 1.

This had implications in turn for how we could identify the impact of what the school was doing. A very detailed articulation of the theory of change would have enabled us to set out sequences of specific changes that might be anticipated to result from specific actions. We could then have set about devising sensitive instruments for identifying and (where appropriate) quantifying such changes. As it was, our more ‘broad brush’ articulation meant that the theory of change acted less as a detailed guide than as a sensitising device, alerting us to the *kinds of* changes that might result from particular *kinds of* action. So, for instance, we agreed with school leaders that the strand of actions directed towards community engagement with learning might be expected to produce (in rough sequence) increased adult-school contact, increased adult participation in learning activities, raised adult aspirations, and re-entry into employment.

We then sought evidence that some or all of these changes were happening. Often, the most illuminating evidence was anecdotal in the sense that it took the form of accounts of and by individuals about how they had been affected by the school’s full service provision. We then set about contextualising these accounts by trying to establish how typical they were of other people’s experiences, who did not access the school’s provision or respond in this way, and what might have happened if the school had not developed this provision. In this way, the process was more exploratory than we had imagined it might be. Rather than demonstrating and quantifying changes

conclusively, we found ourselves gathering indicative evidence on which we could construct a more or less convincing account of the way the school's work was impacting on local people, and of the outcomes that might be anticipated in the long term. Full details and discussions of our findings can be found in the interim year two report (Cummings et al. 2006) and the final report (Cummings et al 2007).

Some reflections

In making our initial decision to use a theory of change approach, we assumed a certain interchangeability between classical outcomes evaluations and the methodology we were about to use. Although we were convinced that simply dropping a limited range of end-point outcomes measures onto the initiative was largely futile, we believed that a theory of change approach would allow us to achieve similar purposes through more context-sensitive means. However, as the brief account we have just provided indicates, the practice of theory of change evaluation – in our case, at least – began to look somewhat different. On reflection, it seems to us that what we were doing embodied some fundamental shifts from the more familiar approaches to outcomes evaluation in education:

1. Shifting the value of what is evaluated

Using a theory of change approach had three significant – and to some extent, unexpected – impacts on what counted as valued outcomes. First, in a context where much of the focus of policy and policy evaluation is on the attainment outcomes privileged by government, the approach required us to focus on the outcomes that were valued by actors in and around schools. For the most part, these were head teachers and school extended service coordinators, who saw themselves as front-line

professionals attempting to make a difference to the lives of children and young people in areas of significant disadvantage. The outcomes that concerned them certainly included attainment outcomes and, arguably, drew heavily on dominant policy discourses (Cummings, Dyson and Todd 2007). Nonetheless, they were by no means confined to attainment, nor, indeed, to outcomes for students. Instead, they embraced a wide range of outcomes for students, families and communities, including, as we have seen, engagement with learning, high aspirations, progression into stable employment, self-esteem, community leadership capacity and community cohesion.

Second, as we have seen, the attempt to establish chains of intermediate changes foregrounded what happened to individuals in a way which we had not entirely anticipated, and which made the focus of our evaluation very different from a more traditional focus on population-level outcomes. This was partly because school leaders found it relatively easy to direct us towards individual students and adults who had dramatic stories to tell about how the school's work had changed their lives (which is why, of course, it was important for us to contextualise these stories). However, it was also because it was often in the lives of individuals (or, equally often, the lives of families) that the complex interventions of schools' full service provision engaged with the complex sets of problems that provision was designed to address. So, our evidence base was full of cases where the school had made available multiple interventions to tackle the multiple problems faced by an individual, where the interventions had changed as the individual's situation had changed, and where, therefore, it was possible to trace a long-term individual trajectory in response to the school's work.

As a result, we found ourselves able to identify outcomes and beneficiaries that might have slipped below the radar of more conventional evaluation designs. Whereas, for instance, the evaluation of the near-equivalent New Community Schools initiative in Scotland was able to find little evidence for an impact on learning outcomes from its analysis of school and student performance data (Sammons et al. 2003), we were able to identify multiple cases of individual students where there was evidence of higher levels of attainment as a result of FSES provision than might otherwise be the case. Similarly, when we sought for evidence of area change through an analysis of neighbourhood statistics, we found nothing. However, when we put together case studies of individual community members, accounts from individuals of how their networks of families and friends were changing, and reports from community professionals of what their observations suggested, we began to see at least the first stirrings of area-level impact.

It is arguable, of course, that the evaluation approach we chose set the bar too low in terms of acceptable evidence, and that more standard methods have failed to find outcomes for the simple reason that there are none (Rees, Power, and Taylor 2007). However, it is, we suggest, equally possible that different approaches to evaluation illuminate or conceal different outcomes for different groups, and that what looks like an absence of outcomes may simply be a methodological insensitivity to the outcomes that are there.

2. Shifting the focus from success and failure to potential

A second unexpected shift was from a conventional focus on the success or failure of the initiative to an assessment of its potential. In more familiar evaluation approaches, outcomes are either evident or they are not, and the success or failure of the initiative can be judged on this basis. As we have seen, however, theory of change approaches contain a predictive element. The question they seek to answer is not simply what outcomes have already materialised, but what outcomes might be anticipated – and this question cannot be answered without building in assumptions about how the current situation will develop in future. Such assumptions, can, of course, contain a range of possibilities, from best to worst case scenarios.

So, in our evaluation, we identified schools where the FSES approach was plainly struggling – but at the same time we found others where things were going well. Likewise, whatever the overall impact of full service provision, leaders in these schools were typically able to point us towards students, families and community members where beneficial impacts were clear. Traditionally, evaluators are often wary of being steered towards such localised success stories, on the grounds that they do not reflect the overall picture of the initiative. In our case, however, we found such stories highly significant. Whilst they were not trustworthy as guides to what the initiative as a whole – or even the work of the school as a whole – was achieving, they told us a great deal about what might be achieved under particular circumstances. In effect, we found ourselves having less to say about the success or failure of the FSES initiative as such, and more to say about the potential of the approach to school roles that was emerging in response to that initiative (see Cummings et al. 2007: 13ff for a fuller discussion).

The danger of presenting an over-optimistic account of an initiative – of treating the exceptional as typical – is, of course, self-evident. By the same token, however, it is unrealistic to suppose that a multi-dimensional, weakly-specified and locally variable initiative such as the one we were evaluating could be uniformly successful. Negative judgements, particularly if made early in the process, may be technically accurate, but may seriously underestimate the potential of such an initiative to have positive impacts in the long term if (and this is an important proviso) it is sustained under reasonably favourable conditions. Arguably, therefore, the identification of such potential and the specification of those conditions is a much more productive outcome from the evaluation process.

3. Shifting the roles of actor and evaluator

In the tradition of formative evaluation, it was always our intention that our work would be of use to leaders of FSESs in developing their work. However, although formative evaluation, as usually understood, changes the point at which evaluation findings are made available and therefore the way in which they can be used by actors in an initiative, it does not in itself change the roles of actor (as the one who plans and does) and evaluator (as the one who identifies successes and failures). This might have been the case in our evaluation if school leaders had been able to articulate, with minimal prompting, a coherent theory of change. However, as other evaluators have found (Spicer and Smith 2008), actors in complex initiatives are not necessarily in a position to do this. In practice, as we have seen, we spent much time working with these leaders effectively to co-construct their theories, so that our work with them was itself an intervention in the situation. As FSES leaders frequently told us, it helped them think through what they were doing and why.

As the initiatives unfolded, this process was extended in two ways. First, because the evaluation took the form of testing leaders' theories of change, the feeding back of early findings was formative not only in telling leaders whether their initiatives was 'working', but in raising questions about its underpinning theory. In other words, it had the potential – to put it no more strongly – of generating a form of 'double loop learning' (Argyris and Schon 1978, 1996) in which both practice and the assumptions underlying practice are problematised and developed. Second, although we could in principle have undertaken all of the data collection ourselves, in practice the focus on relatively small, intermediate changes identifiable in the course of daily practice meant that practitioners themselves were sometimes best placed to gather data. As a consequence, they became participants in the evaluation process rather than simply recipients of its findings.

In these ways, the border between evaluation and developmental consultancy in this evaluation was one that was frequently crossed, while the roles of actors and evaluators became, if not interchangeable, at least much closer than is usual. In fact, theory of change evaluation has many similarities to theory based approaches to planning, such as logic modelling (W.F. Kellogg Foundation 2004), and to peace education (Ashton 2007) and it is no surprise, therefore, that it has been used as a consultancy tool in the design stage of initiatives (Connell and Klem 2000).

Flexible evaluation for complex initiatives

We would not contest the charge that the shifts we experienced in using a theory of change approach were to some extent the consequence of our failure to appreciate

fully at the outset what such an approach might involve, and that we allowed ourselves to drift into a more exploratory and open-ended approach than was strictly necessary. However, we do not think this tells the whole story. We also believe that the more holistic approach to children's services that is emerging in England, and the consequent extension and redefinition of the role of schools that is happening in many countries, demand different forms of evaluation from those that have tended to be dominant in recent years.

We say this for a number of reasons. First, as we have seen, the nature of interventions is changing in response to this more holistic approach. Single-strand interventions with clearly specified outcomes are being supplemented (if not replaced) by multi-strand interventions aiming at less well-specified, and more wide-ranging outcomes. Second, the complexity of these interventions and of the contexts in which they are to be implemented means that the capacity of policy makers at the centre to design them in fine detail is severely limited. Third, and as a consequence of this, the role of local actors changes. Rather than being simply implementers of centrally-designed initiatives, they become policy makers in their own right (Ozga 2000). As the Government itself has recognised:

Holistic government in particular places cannot be imposed top-down from a distance. If frameworks for co-operation are to be effective, they need to be more than lists of externally imposed priorities. They must also reflect the whole needs of communities and the priorities of local people. Joined up working must create room for personal initiative and creativity.

(National Strategy for Neighbourhood Renewal 2000: par. 1.17)

Fourth, this cocktail of complexity, indeterminacy and local ‘creativity’ means that initiatives are, at local level, likely to be variable in form and aim, and, moreover, to be subject to variation over time as contexts, personnel, and the understandings of their leaders change.

It follows that evaluations have to engage with these high levels of complexity, indeterminacy and variability. In doing so, they may have to accept that the model of evaluator as detached observer passing authoritative judgements on success and failure is no longer entirely appropriate. Certainly, simply dropping a set of externally-formulated outcome measures onto these complex situations – at least, if that is *all* the evaluation does – is likely not only to tell us little about the initiative, but also to lead to promising practices being damned simply because of the insensitivity of evaluation methods.

We would not wish to argue that the theory of change approach is the perfect solution in this situation. Despite its relative youth in its current form, it has already come under attack for, amongst other things, being too cumbersome, too linear, and too apolitical (see, for instance, Mackenzie and Blamey 2005; Davies 2005; Davies 2004; Edwards et al, 2006). Nor is it the only solution that is available. The national evaluators of the Children’s Fund – an initiative recognisably from the same stable as FSESs and Sure Start – for instance, found it necessary to supplement their use of theory of change with an approach drawing on activity theory (Edwards et al. 2006: 247ff), and there may be something to be said for combined approaches of this kind. The issue at this juncture, however, is that what seems to be important is not so much

the relative merits of different approaches as the extent to which evaluators will be able to keep pace with the changing realities of schools' expanding roles.

References

- Ashton, C. V. 2007. Using Theory of Change to Enhance Peace Education Evaluation. *Conflict Resolution Quarterly* 25, no. 1: 39-53.
- Argyris, C., and D. A. Schon. 1978. *Organizational learning*. Reading, M A: Addison-Wesley.
- . 1996. *Organizational learning II: theory, method and practice*. Reading, M A: Addison-Wesley.
- Ball, M. 1998. *School Inclusion: The school, the family and the community*. York: Joseph Rowntree Foundation.
- Ball, S.J. 2008. *The Education debate: policy and politics in the Twenty-First Century*. Bristol: The Policy Press.
- Barnard, M, E. , C. Becker, N. Creegan, K. Day, E. Devitt, L. Fuller, H. Lee, S. Neil, S. Purdon, and H. Ranns. 2009. Evaluation of the National Healthy Schools Programme: interim report. London: National Centre for Social Research for the Department of Health.
- Belsky, J., J. Barnes, and E. Melhuish, eds. 2007. *The National Evaluation of Sure Start: does area-based early intervention work?* Bristol: The Policy Press.

- Blair, T. 1999. Speech by the Prime Minister Tony Blair about Education Action Zones - 15 January 1999. 10 Downing Street 1999. <http://www.number-10.gov.uk/output/Page1172.asp>.
- Blank, M., A. Melaville, and B. Shah. 2003. Making the difference: research and practice in community schools Washington DC: Coalition for Community Schools, Institute for Educational Leadership.
- Brooke-Smith, R. *Leading Learners, Leading Schools*. London: Routledge-Falmer.
- Chapman, C., and H. Gunter. 2009. A decade of New Labour reform of education. In *Radical reforms: perspectives on an era of educational change.*, edited by C. Chapman and H. Gunter. London: Routledge.
- Clark, Helene, and Clareann Grimaldi. 2005. Evaluation of Children's Aid Society Community Schools. In *Community schools in action: lessons from a decade of practice*, edited by J. G. Dryfoos, J. Quinn and C. Barkin. Oxford: Oxford University Press.
- Connell, J., and A. M. Klem. 2000. You can get there from here: using a theory of change approach to plan urban education reform. *Journal of Educational and Psychological Consulting* 11, no 1. : 93-120.
- Connell, J. P., and A. C. Kubisch. 1998. Applying a theory of change approach to the evaluation of comprehensive community initiatives: progress, prospects and problems. In *New approaches to evaluating community initiatives. volume 2: theory, measurement and analysis.*, edited by K. Fulbright-Anderson, A. C. Kubisch and J. P. Connell. Queenstown: The Aspen Institute.
- Coote, A., J. Allen, and D. Woodhead. 2004. *Finding out what works: building knowledge about complex community-based initiatives*. London: The King's Fund.

- Cummings, C., A. Dyson, D. Muijs, I. Papps, P. Pearson, C. Raffo, L. Tiplady, and L. Todd. 2007. Evaluation of the Full Service Extended Schools Initiative: Final Report. London: DfES Research Report 852.
- Cummings, C., A. Dyson, I. Papps, D. Pearson, C. Raffo, and L. Todd. 2005. Evaluation of the Full Service Extended Schools Project: End of First Year Report. London: DfES.
- Cummings, C., A. Dyson, I. Papps, P. Pearson, C. Raffo, L. Tiplady, and L. Todd. 2006. Evaluation of the full service extended schools initiative, second year: thematic papers. London: DfES.
- Cummings, C., A. Dyson, and L. Todd. 2007. Towards extended schools? How education and other professionals understand community-oriented schooling. *Children and Society* 21:189-200.
- Davies, R.J. . 2004. Scale, complexity and the representation of theories of change. *Evaluation* 10, no.1: 101-121.
- . 2005. Scale, complexity and the representation of theories of change part II. *Evaluation*. *Evaluation* 11, no. 2: 133-149.
- DCSF. 2009. Schools Analysis and Research Division, Department for Children Schools and Families. 2009. Deprivation and education: the evidence on pupils in England, Foundation Stage to Key Stage 4. London: DCSF.
- DfEE. 1999. Schools Plus: building learning communities. Improving the educational chances of children and young people from disadvantaged areas: a report from the Schools Plus Policy Action Team 11. London: DfEE (Department for Education and Employment).
- . 1999. Sure Start: Making a difference for children and families. London: DfEE (Department for Education and Employment).

- DfES. 2003. *Every Child Matters*. London: HMSO.
- . 2003. *Full service extended schools: requirements and specifications*. London: DfES.
- . 2003. *Full-service extended schools planning documents*. London: DfES.
- . 2005. *Extended schools: access to opportunities and services for all. A prospectus*. London: DfES.
- Dryfoos, J. 1994. *Full-Service Schools: A Revolution in Health and Social Services for Children, Youth, and Families*. San Francisco: Jossey-Bass.
- Dryfoos, J., and S. Maguire. 2002. *Inside full-service community schools*. Thousand Oaks, CA: Corwin Press.
- Edwards, A., M. Barnes, I. Plewis, K. Morris, and et al. 2006. *Working to prevent the social exclusion of children & young people: final lessons from the National Evaluation of the Children's Fund. Research Report RR734*. London: DfES.
- Foundation., W.K.Kellogg. 2004. *Logic model development guide*. W.K. Kellogg Foundation.
- Geddes, M. 2006. *National Evaluation of Local Strategic Partnerships: theory of change issues paper*. London: DCLG.
- Gorard, S. 2005. Academies as the 'future of schooling': is this an evidence-based policy? . *Journal of Education Policy* 20, no. 3: 369 - 377.
- Henderson, A. T., and K. L. Mapp. 2002. *A new wave of evidence: the impact of school, family, and community connections on student achievement: annual synthesis 2002*. Austin, Texas: Southwest Educational Development Laboratory.

- Humphrey, N., K. Kalambouka, J. Bolton, A. Lendrum, M. Wigelsworth, C. Lennie, and P. Farrell. 2008. Primary Social and Emotional Aspects of Learning (SEAL): evaluation of small group work. London: DCSF.
- Johnson, M. 2008. Ofsted social targets misunderstand schools. . *EducationGuardian.co.uk* 30 April 2008.
- Keyes, M. C., and G. Soleil. 2001. *School-community connections: a literature review*. . Charleston, WV: AEL, Inc.
- Mackenzie, M., and A. Blamey. 2005. The practice and the theory: lessons from the application of a theories of change approach. *Evaluation* 11 (2):151-168.
- Morrison, K. 2002. *School Leadership and Complexity Theory*. London: Routledge Falmer.
- National Audit Office. 2007. The Academies Programme. London: HMSO.
- National Evaluation of Sure Start. No date. *National Evaluation of Sure Start - methodology report executive summary*. Available from <http://www.ness.bbk.ac.uk/documents/Methodology.pdf>.
- National Strategy for Neighbourhood Renewal. 2000. *Report of Policy Action Team 17: Joining it up locally*. London: London: Department of the Environment, Transport and the Regions.
- NFER. 2007. National Evaluation of Excellence in Cities 2002 - 2006. London: DCSF.
- Ozga, J. 2000. *Policy research in educational settings: contested terrain*. Buckingham: Open University Press.
- PriceWaterhouseCoopers LLP. 2007. Academies evaluation: 4th annual report. London: DCSF.

- Rees, G., S. Power, and C. Taylor. 2007. The governance of educational inequalities: the limits of area-based initiatives. . *Journal of Comparative Policy Analysis* 9, no. 3: 261-274.
- Sammons, P., S. Power, K. Elliot, P. Robertson, C. Campbell, and G. Whitty. 2003. New Community Schools in Scotland. Final report. National evaluation of the pilot phase. . London: Institute of Education, University of London.
- Schools Analysis and Research Division, Department for Children Schools and Families. 2009. *Deprivation and education: the evidence on pupils in England, Foundation Stage to Key Stage 4*. London: DCSF.
- Spicer, N., and P. Smith. 2008. Evaluating complex, area-based initiatives in a context of change: the experience of the Children's Fund initiative. *Evaluation* 14:75-90.
- Stame, N. 2004. Theory-based evaluation and types of complexity. *Evaluation* 10, no. 1: 58-76.
- Sullivan, H., M. Barnes, and E. Matka. 2002. Building collaborative capacity through theories of change: early lessons from the evaluation of Health Action Zones in England. *Evaluation* 8, no. 1: 205–26.
- Szirom, Tricia , R. Jaffe and D. MacKenzie (Strategic Partners in association with the Centre for Youth Affairs and Development). 2001. *National evaluation report. Full Service Schools Program 1999 and 2000*. Canberra: Commonwealth Department of Education, Training and Youth Affairs.
- Toynbee, Polly. 2005. Comment. We must hold our nerve and support deprived children. Sure Start remains one of the government's best achievements, but the new centres need more professional backup.
- <http://www.guardian.co.uk/education/2005/sep/13/earlyyearseducation.uk>.

- Weiss, C. H. . 1995. Nothing as practical as a good theory: exploring theory-based evaluation for comprehensive community initiatives for children and families. In *New approaches to evaluating community initiatives: concepts, methods and contexts*, edited by J. Connell, A. C. Kubisch, L. B. Schorr and C. H. Weiss. Washington DC: The Aspen Institute.
- Weitzman, B. C. . 2002. Integrating a comparison group design into a theory of change evaluation. *American Journal of Evaluation* 23, no. 1: 371-385.
- Whalen, S. 2002. Report of the evaluation of the Polk Bros. Foundation's Full Service Schools Initiative. Executive summary. Chicago, Illinois: Chapin Hall Center for Children at The University of Chicago for the Polk Bros Foundation.
- Wilkin, A., K. Kinder, R. White, M. Atkinson, and P. Doherty. 2003. Towards the Development of Extended Schools. Research Report 408. London: NFER/ DfES.
- Wilkin, A., R. White, and K. Kinder. 2003. Towards Extended Schools: A Literature Review. Research Report 432. London: NFER/ DfES.
- W.K. Kellogg Foundation. 2004. *Logic model development guide*. Battle Creek, MI: W.K. Kellogg Foundation.
- Wyse, D. , E. McCreery, and H Torrance. 2008. The trajectory and impact of national reform: curriculum and assessment in English primary schools (Primary Review Research Survey 3/2). Cambridge: University of Cambridge Faculty of Education.

Figure 1. Simplified diagram of a school's theory of change



