The challenges to effective outcome evaluation of a national, multi-agency initiative: The experience of Sure Start
Nigel Lloyd and Louise Harrington
Evaluation 2012 18: 93
DOI: 10.1177/1356389011429631

The online version of this article can be found at: http://evi.sagepub.com/content/18/1/93
The challenges to effective outcome evaluation of a national, multi-agency initiative: The experience of Sure Start

Nigel Lloyd
NLH Partnership Ltd, UK

Louise Harrington
NLH Partnership Ltd, UK

Abstract
This article uses the example of Sure Start, a national initiative introduced by the UK Government in the late 1990s, as a case study to explore the reasons why large-scale, complex, national initiatives often fail to adequately evidence the impact of their work. The authors explore a range of structural, cultural, methodological and practical factors that have acted to inhibit effective evaluation of the impacts of Sure Start. They argue that the potential exists for more effective and efficient evaluation of the outcomes of complex community initiatives such as Sure Start, if factors such as those outlined within the article are adequately addressed. This article provides important learning for those funding, commissioning and delivering complex community initiatives.

Keywords
complex community initiatives, factors inhibiting effective evaluation, local and national evaluation, outcome evaluation, Sure Start

Introduction
Evaluating complex community initiatives (CCIs) such as Sure Start can be difficult (Barrett, 2007; Kemp et al., 2008; Spicer and Smith, 2008). This article takes the Sure Start initiative and, in particular, the evaluation experience of Sure Start Local Programmes (SSLPs) as a case study to investigate this.
The article begins with an overview of Sure Start, and its evaluation. It then outlines a number of factors that have made it difficult for the impacts of Sure Start to be assessed effectively. The article explores the factors that inhibited effective outcomes evaluation of SSLPs and argues that exploration of these factors provides important learning for those attempting to evaluate the impacts of large-scale national initiatives. It also argues that the potential exists for more effective and efficient evaluation of the outcomes of Sure Start Children’s Centres and other national initiatives, and that there are a number of steps that Sure Start and similar initiatives can take in order to better assess their effectiveness.

**Background to Sure Start**

The Sure Start initiative was introduced by the UK Government in the late 1990s, as a result of the Cross-Departmental Review of Provision for Young Children, which examined a range of services provided for children aged 0–7. The Review found that many young children and their families in the UK, and particularly those living in certain ‘disadvantaged’ areas, were not receiving the support they required, and that evidence suggested a need for greater investment in the early years of children’s lives (HM Treasury, 2008), in order to positively influence later life chances and developmental, health, educational and social outcomes. The Review concluded that a range of accessible, local, integrated and coordinated services should be in place to provide support to young children and their families, and that these services should be planned and delivered through joint working between all relevant organizations.

The integrated, area-based approach to service delivery recommended by the Review was implemented through the introduction of the Sure Start programme in 1998. Sure Start focused on providing enhanced services for families with children aged under four. It had a clear remit to target services at particularly ‘deprived’ local communities, and was targeted at the 20 percent most deprived areas in England (DfEE, 1999).

During the period 1999 to 2003, 524 Sure Start Local Programmes (SSLPs) were introduced across England, providing services for families living within a specified postcode ‘reach’ area. Local authorities, Primary Care Trusts, and national children’s charities were usually the bodies responsible for developing and implementing SSLPs within communities, in partnership with other local agencies, voluntary organizations and the local community. Collaboration and joint working among local partners and agencies was central to the Sure Start ethos, as were the principles of community and parental involvement in Sure Start management and decision-making (Ball, 2002; DfEE, 1999).

Specific targets and objectives for SSLPs changed over time and local priorities varied from programme to programme. The work of SSLPs was targeted at:

- Improving learning;
- Improving children’s health;
- Improving social and emotional development;
- Strengthening families and communities. (Sure Start Unit, 2001)

Different SSLPs took different approaches to addressing these targets and objectives, delivering services tailored to meet the needs of families living within their catchment area. Innovation and creativity in terms of service coordination and delivery were encouraged. The list of services and activities delivered nationally by SSLPs is vast, but typical services included: parents and toddler
groups, parenting courses and support, employment and training for parents, groups for teenage parents, in-depth family support work, support for children with additional needs and their families, breastfeeding peer support, outreach and home-visiting support, speech and language development support, and diet and nutritional advice.

**Recent development of the Sure Start initiative**

The New Labour Government reinforced its commitment to improving the lives of young children and their families through its decision to implement a national programme of Sure Start Children’s Centres (SSCCs), aimed at families with children aged under five. This represented a ‘roll-out’ of the Sure Start initiative across the whole of England, rather than the previous focus on particularly ‘deprived’ communities, and involved the vast majority of previous SSLPs becoming SSCCs, as well as new SSCCs being created to meet the Government’s goal of having ‘a Children’s Centre for every community’. The roll-out of Children’s Centres began in 2004.

The current coalition Government has stated its commitment to Sure Start Children’s Centres, although it proposes a move away from the ‘universal’ Children’s Centre model pursued by the New Labour Government, and towards the targeting of services at the ‘neediest’ and most vulnerable children and families (Asthana, 2010).

In addition, there has been a great amount of recent policy focus on the importance of ‘early intervention’ and the provision of adequate support for babies and young children and their families, of the type championed by Sure Start. This has included analysis of the potential financial savings that might be made through increased investment in the early years, improving the early development, experiences, and life chances of children, and thereby reducing the need for costly expenditure on interventions to ‘undo the damage’ later on in children’s lives (see e.g. Allen, 2011; Field, 2010). Recent influential reviews of evidence such as the Graham Allen Review (Allen, 2011) highlight the important role of Children’s Centres in delivering early years services for children and families.

**The evaluation of Sure Start**


The NESS ‘Impact’ module was focused on investigating the impacts of the Sure Start initiative nationally. The Impact module’s methodology included cross-sectional and longitudinal study of more than 9000 children and their families living within SSLP areas. The NESS Impact study utilized a range of quantitative measures and tests, to explore factors such as child cognitive ability, maternal well-being, and parenting/family functioning (for example, ‘negative parenting’ and ‘home chaos’) (NESS, 2005).

The second ‘level’ of evaluation was ‘local evaluation’. Government guidance required each SSLP to conduct a locally-led annual evaluation of its services (Sure Start Unit, 2002), and to submit an annual, programme-wide evaluation report to the Government’s Sure Start Unit (SSU), the cross-departmental unit that co-ordinated the delivery of Sure Start.
Programmes had a great deal of autonomy regarding what they evaluated and when, and adopted varied approaches to local evaluation. Programmes typically employed one of the following approaches to their local evaluation:

- **Internal**: The evaluation was conducted entirely from ‘within’ the programme, often by a staff member for whom evaluation was a primary or key role;
- **External**: The majority of evaluation work was commissioned out to an ‘external’ evaluator or evaluation team from a research consultancy or university;
- **Mixed**: Some evaluation work was conducted internally, with some external evaluation work commissioned out (often evaluations of specific services or topics), as and when required. (Harrington et al., 2005)

To support SSLPs with their local evaluation, the NESS ‘Support for Local Programmes’ module was tasked with providing hands-on, on-going evaluation guidance, support and expertise to every SSLP in England. Each of the nine English Government Office regions was allocated a NESS Support ‘Regional Support Officer’ to provide support and guidance with evaluation issues to all SSLPs within that region.

**The authors’ involvement with Sure Start**

The authors of this article have been involved with Sure Start since 2002. We were members of the NESS ‘Support for Local Programmes’ teams, first as Regional Support Officers and later as National Coordinators, and provided research and evaluation support and guidance to all 153 SSLPs in the London, East, and South East regions of England. Whist at NESS we wrote and contributed to a number of evaluation-related documents (see e.g. Harrington et al., 2005; Lloyd and Rafferty, 2006; Lloyd et al., 2003; Moran and Harrington, 2006).

Since setting up NLH Partnership we have continued to work closely with Sure Start and conducted numerous Sure Start Children’s Centre research, evaluation, performance management, and development projects, for local Centres, and local, regional and national Government.

This article draws on our first-hand experience and knowledge of supporting, coordinating and conducting Sure Start evaluations at a national and local level.

**The need for good evidence**

The effective evaluation of publicly funded Government initiatives such as Sure Start is important for a number of reasons. In particular, the recent changes in UK Government, and the extensive public-sector spending cuts, have made the need for evidence of the impacts and outcomes of national, publicly funded initiatives more acute. Impact and outcome evidence can help to safeguard effective services by providing valuable evidence of the efficacy of initiatives and the need for continued funding. Such evidence can also assist with the drive for austerity and efficiency by helping to highlight the most effective and cost-effective way of delivering positive impacts, and developing ways of gaining the same or greater impacts without increased expenditure.

The cost of initiatives such as Sure Start provides a further driver for good quality outcomes evidence. Sure Start was a major national initiative, costing in excess of £450,000,000 between 1998 and 2000 alone, and in excess of £3 billion between 2004 and 2008 (National Audit Office, 2006). Despite this expenditure, since the start of Sure Start there has been a lack of evidence...
related to the impacts and outcomes of Sure Start for children and families, and the difference that the initiative makes to families and communities.

The report of the first phase of the NESS Impact evaluation in 2005 highlighted only limited and small effects of SSLPs, with some negative effects of SSLPs also highlighted (NESS, 2005). The second phase Impact evaluation was more promising; however, even at this stage NESS reported relatively modest findings about the impacts of SSLPs (NESS, 2008).

The latest study from the National Evaluation of Sure Start (NESS, 2010) reported a number of key impacts of SSLPs for children and their families. Among the study’s key findings were that children in SSLP areas had lower BMI and experienced better physical health than those in non-SSLP areas, and that there were a number of positive effects of Sure Start on mothers’ well-being and the ways in which families ‘function’. Overall though, these NESS findings again showed relatively limited effects of Sure Start, with no positive SSLP effects for six variables, two negative SSLP effects, and a large number of non-effects of Sure Start, particularly with regard to children’s development (NESS, 2010).

It has, then, been difficult to demonstrate the impacts of Sure Start at a national, NESS level. At the same time, those of us who have worked closely with SSLPs and Children’s Centres over the past few years are likely to be able to recount numerous examples of Sure Start’s positive impact upon the lives of local community members, in some cases transforming the lives of children and families. The ‘on-the-ground’ experience – particularly where programmes have been around since the early days of Sure Start and have had close to 10 years to embed themselves within communities – is often of an initiative that has had huge impacts for many families and that has great potential to bring about positive outcomes (Harrington and Lloyd, 2010).

Factors inhibiting the effective assessment of Sure Start outcomes

Structural

Lack of integration between the local and national levels of evaluation. The methodology for measuring the impacts of Sure Start included an explicit distinction between the local and national levels. National-level impact evaluation conducted by NESS aimed to assess the overall impact of Sure Start, and individual SSLP evaluations focused on individual programmes evaluating their own work, largely in isolation from other SSLPs.

There was little formal recognition of the potential value to the overall national evaluation (of the Sure Start initiative) of incorporating findings from local evaluation. NESS did attempt to bring together findings from local SSLPs, through an online database of programmes’ evaluation reports and through the production of ‘synthesis’ documents that amalgamated findings related to different evaluation themes (see e.g. Lloyd and Rafferty, 2006; Moran and Harrington, 2006). However, these documented the local evaluation work conducted across the country but did not attempt to integrate local-level data into the national evaluation’s body of data, or to use what local evaluation expertise existed to inform the national evaluation.

Overall then, there was a lack of integration between the national and local evaluation levels, and no clear mechanism for bringing together evidence from local evaluations in a coherent way, so that locally collected impact information could feed into national-level assessment of impact. Such an evaluation methodology represented a missed opportunity to integrate local and national levels, and allow the two to better inform each other.
The only significant attempt to integrate the local and national evaluation levels came through the implementation of the Sure Start Language Measure (SSLM), which was designed by a team at City University to measure children’s early speech and language development. The measure was administered at a local level, with data being passed for analysis at the national level, as well as being available for local analysis (Harris et al., 2005).

**The focus on evaluating ‘process’**. A key reason for the lack of SSLP outcomes evaluation evidence was that most SSLPs focused their evaluation efforts on evaluating ‘process’. Evaluation requirements were laid out in Sure Start Unit guidance (Sure Start Unit, 2002). Initial guidance suggested that SSLPs should focus their evaluation efforts on ‘process’ issues (e.g. exploring joint working or decision-making processes), rather than ‘impact’. There is clearly logic to this, as the impacts of initiatives such as Sure Start cannot be measured at the outset. However, the initial focus on ‘process’ evaluation had meant that SSLPs had developed systems, processes and resources focused around measuring process rather than impact, and as such, evaluation had become synonymous with measuring ‘process’. Thus, when Sure Start Unit guidance later shifted towards greater encouragement of outcome evaluation, it was difficult for programmes to react adequately to the new drive for such evidence.

The initial focus on evaluating process meant that opportunities were often missed to implement appropriate outcomes-focused evaluation methodologies, and collect the necessary baseline and other data.

**Lack of coordination and integration of different local evaluations.** SSLP local evaluation was characterized by individual programmes conducting their own, individual programme evaluations, with little link between these evaluations and other local evaluations or the national evaluation of Sure Start. Early SSU evaluation guidance instructed SSLPs to take this individualized approach to local evaluation and discouraged joint evaluation, for example between programmes that bordered each other or that were within the same or neighbouring local authority areas (Sure Start Unit, 2002). There was therefore a lack of coordination and integration between the local evaluations of different SSLPs.

This individualized approach to local evaluation suggested that despite opportunities provided through regional events and online discussion forums, for local programmes to share concerns and issues, local evaluation was to be thought of as primarily an individual programme endeavour.

There was some commonality in terms of the evaluation themes that different SSLPs focused on. However, there was a lack of formal systems and protocols to ensure the coordination of evaluation themes or approaches across SSLPs. This made it difficult for consistent approaches to local evaluation to be developed, or for impact findings to be consolidated across programmes.

A key reason for the SSU’s steer towards individualized programme evaluations may have been a desire for the SSLP evaluation ethos to mirror the ethos of the SSLP initiative. SSLPs were individual programmes aimed at children and families living within a specified catchment area, with the autonomy to develop, coordinate and deliver services in the way that best suited local need. The individualized approach to evaluation encouraged by the SSU clearly mirrors the individualized ethos of the early Sure Start initiative. Hills (2004) is among those who have drawn attention to this tendency towards evaluations of community-level programmes that reflect the ethos of the programmes themselves.

Hills (2004) also highlights the tendency for the evaluation guidance provided to community-level programmes to steer them towards relatively narrow, programme-focused evaluation work, and away from evaluation work that generates knowledge and evidence with a broader relevance.
for the policy or research communities. This was certainly true in the case of Sure Start. SSU guidance steered local programmes towards conducting local ‘process’ evaluation work, and away from taking broader approaches, such as conducting cross-programme, thematic, impact evaluation work.

**Limited good quality local evaluation impacts and outcomes evidence.** If local evaluation evidence was to be of practical use in helping to assess the overall impact of SSLPs, it was important that local evaluation evidence was of high quality. The NESS ‘Support for Local Programmes’ module provided each SSLP with a NESS researcher to offer advice and guidance on evaluation-related issues, as well as other support such as workshops and guidance documents. While some programmes took advantage of this support, others did not, and this variability in terms of engagement with available support, contributed to difficulties for some individual programmes in producing good quality impact evidence.

In addition, where local evaluations were commissioned out to an external organization, the resulting evaluation work, whether focusing on processes or impacts, was often of relatively poor quality and there was generally a lack of good-quality, locally researched evidence about the outcomes and impacts of SSLPs.

**Practical**

**Focus on the most pressing issues.** Where time and resources are limited, it makes sense for projects to focus on their most pressing work, and those tasks that have explicit requirements and deadlines. This happened with Sure Start. Understandably, many SSLPs initially focused their attention on the job of building local partnerships, commissioning local services, sourcing venues and buildings, and delivering services for children and families, rather than on conducting evaluations.

There were requirements from the Sure Start Unit for programmes to regularly collect and submit monitoring data. Evaluation requirements were much less rigid, with poorly enforced, annual deadlines and requirements. Monitoring data thereby became the focus of many programmes’ monitoring and evaluation work. Monitoring, evaluation and performance management efforts were often focused on this monitoring requirement, rather than on conducting evaluation work.

**Difficulty engaging key partners.** Effective evaluation of a complex, multi-agency, area-based initiative such as Sure Start involves the input and commitment of the range of partner agencies and organizations involved in the local delivery of services for families with young children, such as health visiting and midwifery teams, social care, and family support agencies (Harrington and Lloyd, 2008; Lloyd and Harrington, 2008). Each partner had a key role to play in the effective evaluation of the services they were involved with, and of the programme overall. Partner agencies’ and professionals’ understanding of their role in the evaluation process was essential if evaluation was to prove effective. Without such understanding and partner involvement, it was impossible for SSLPs to gain the necessary information that they needed to effectively evaluate the impacts of their services.

However, encouraging the meaningful involvement of partner organizations, at either a strategic or operational level, was an enduring challenge for SSLPs and continues to provide difficulties for current Children’s Centre partnerships. Where strategic commitment to joint and partnership working was evident, this commitment sometimes failed to result in effective operational, on-the-ground joint-working, and vice versa.
One reason for this was that the scope of Sure Start partnerships meant that a wide range of partners with differing perceptions and experiences of evaluation had been brought together under the Sure Start banner. Building a commitment to evaluation among these disparate groups – albeit ostensibly working towards shared objectives – was a major challenge, and often resulted in partners having a lack of practical ‘involvement’ in evaluation and the data-collection processes that were central to it.

A simple but common example was where SSLPs commissioned a provider to deliver a certain service or activity. It was common for some of these providers to fail, or refuse, to collect simple information on attendance, let alone be actively involved in the impact evaluation of their services. Some SSLPs attempted to rectify this issue by introducing evaluation requirements into Service Level Agreements (SLAs) between SSLPs and their partner providers. However, this strategic commitment to collecting information and conducting evaluation work often did not translate into practice.

A number of commentators have highlighted the barriers that exist to encouraging the involvement of key stakeholders in evaluation. Skinner (2004) for example, highlighted barriers such as previous negative experience of evaluation, lack of strategic commitment to evaluation, time pressures, and poor framing of evaluation aims and objectives. Similarly, Donaldson et al. (2002) drew attention to the detrimental impact that fear of evaluation, and more specifically ‘excess evaluation anxiety’, may have on attempts to effectively engage all key stakeholders in the evaluation process. Barriers of the type highlighted by Skinner and Donaldson et al. certainly contributed to the difficulty that some SSLPs found in engaging key partners in local evaluation. In particular, lack of time to participate in evaluation as a result of being short-staffed was typically cited by some partners, such as health visitors, as a reason for their lack of participation in evaluation. Fear of evaluation, which sometimes manifested itself as resistance to participating in evaluation, was also apparent among some SSLP partners. SSLPs then, found difficulty in building a programme-wide and partner-wide commitment to evaluation.

Lack of local evaluation capacity and expertise. Regardless of whether SSLPs took an internal, external or mixed approach to evaluation, there tended to be a lack of knowledge and understanding within SSLPs, about how to effectively evaluate the impacts of Sure Start. Internal evaluators, for example, were rarely highly experienced social researchers when appointed to their posts. External consultants, commissioned by local programmes to offer evaluation expertise, often lacked the technical skills necessary to effectively evaluate programmes’ impacts and outcomes.

This lack of local evaluation capacity and expertise meant that SSLPs struggled to adequately measure the impacts of their work and the outcomes for children and families. McKie (2003) and Hills (2004) are among those who have drawn attention to the lack of evaluation skill and expertise that is sometimes apparent at the local evaluation level of community-based initiatives, and the need for this to be addressed through the provision of specialist support and guidance.

Cultural

The belief that impacts and outcomes can’t be measured locally. Staff and management at many SSLPs subscribed to the view that outcomes either could not be measured at a local level, or that such evaluation was not the remit of Sure Start local evaluation. Different factors contributed to this belief, and in particular, Sure Start Unit guidance that suggested a focus on ‘process’ rather than ‘outcome’, and the preventative nature of Sure Start, which focused on young children and
their families, and the fact that many of the major impacts of Sure Start might not be apparent for many years.

The very nature of Sure Start services, also contributed to the perception that impact and outcome evaluation was difficult or impossible. Many of SSLPs’ early activities were informal ‘interventions’ such as ‘drop-in’ type activities, or other activities and services that were small scale or personally tailored, allowing parents to gain valuable advice, and peer and family support, and increase social capital and reduce social isolation. These were services and activities, the impacts of which programmes found difficult to evaluate in part because it was sometimes difficult to operationalize exactly what ‘intervention’ had been received by the parent, child or family. There was sometimes also a lack of clarity about what the outcomes of the actual service or activity should or might be – a well-documented barrier to effective evaluation (NESS, 2010; Weitzman and Silver, 2003). For such services, outcomes seemed more personal and less tangible.

‘Other-focused’ approach to evaluation. Perhaps the most important cultural and attitudinal factor that inhibited effective impact evaluation was the belief that SSLPs were conducting evaluation primarily for external reasons and for external organizations – that the act of evaluating their services was a problematic requirement that needed to be met, rather than a process conducted for the benefit of the programme and its recipients. Thus, the key driver for evaluation was often seen as ‘external’ – a task undertaken because ‘it was part of the funding requirements’ or ‘because we have to’, rather than because of an understanding of the intrinsic value of both process and outcome evaluation, to service delivery and programme impact.

This external focus also acted to hinder the development of evaluation, with programmes often limiting their work to what they perceived to be the Sure Start Unit requirements. Therefore local evaluations were often reactive rather than proactive. They often focused on the need to produce an evaluation report by the January deadline, rather than on completing an evaluation that would be valuable to the programme, and inform development and an understanding of programme impacts. For many SSLPs, routine impact evaluation was not fully integrated into the overall service-planning process, and evaluation tended not to be incorporated as an integral part of planning and delivering local services.

Methodological

Limited embedding of evaluation within on-going practice and delivery. It was relatively rare for SSLPs to embed good evaluation practice, systems, and processes within on-going service delivery. This added to the difficulty in evaluating impacts and outcomes. The priority for programmes was on setting up and delivering services. This often meant that there was limited focus on how best to evaluate the impacts of those services.

In addition, the precise aims and objectives of services were often poorly defined, making appropriate indicators of change difficult to define and adding difficulty to the task of developing appropriate measures of impact for parents, children, families or communities.

Blamey et al. (2008) highlight the need for evaluation to be built into the delivery plans of CCIs so that a link is identified, at an early stage, between issues that initiatives wish to address, interventions, targets, milestones, and indicators of change. Blamey and colleagues argue that a ‘theory-based’ approach such as this would help those delivering CCIs to focus on how their initiative aims to bring about change, and help with the process of evaluating impacts. Although SSLPs were required to conduct an evaluation, SSLP delivery plans were not required to contain a detailed evaluation plan that embedded evaluation firmly within programme planning.
Such an approach, which would have helped to elevate the profile of evaluation within SSLPs, may also have assisted SSLPs in focusing on impact evaluation at an earlier stage, and encouraged staff and partners to become actively engaged in evaluation processes.

**Lack of a ‘whole-system’ approach to evaluation.** Few SSLPs had developed effective systems for evaluating the impact and outcomes of their services. Where systems were developed, they tended to be piecemeal, focusing perhaps on the evaluation of a few services or activities or an area of work, or dealing only with some aspects of the evaluation process and neglecting others.

Few SSLPs took a ‘whole-system’ approach to evaluation of their programme and developed systems, processes, procedures, protocols and tools to ensure that the impacts of their services and activities could be appropriately evaluated on an on-going basis.

**Towards more effective impact and outcomes evaluation**

We have outlined above, a number of barriers to effective impact and outcomes evaluation of the SSLP initiative. Below, we outline a number of practical solutions for overcoming these barriers and facilitating more effective impact and outcome evaluation of Sure Start and similar initiatives

**Better integration of the national and local evaluations**

The evaluation of SSLPs was characterized by a disconnection between the local and national levels of evaluation. Greater integration of these two levels had the potential to facilitate more effective impact and outcome evaluation.

For instance, a range of impacts of Sure Start for children and families have been captured at a local level (Harrington and Lloyd, 2006, 2007, 2009a; Lloyd and Harrington, 2007), including decreased social isolation and enhanced social networks (Harrington and Lloyd, 2006, 2009a, 2009b), improved parenting skills and knowledge (Harrington and Lloyd, 2006, 2008, 2009a), improvements in parents’ general levels of confidence and their likelihood to engage with local services (Harrington and Lloyd, 2009a; Lloyd and Harrington, 2009), and the development of parents’ employment-related skills and knowledge (Harrington and Lloyd, 2008a, 2009c). For many of the families who engage with Sure Start, such impacts represent life-changing progress and are vital stepping-stones to other longer-term impacts (Harrington and Lloyd, 2009a; Lloyd and Harrington, 2009).

Such local-level impact data provides valuable evidence of the impact of Sure Start. The lack of connection between the local and national levels meant that such ‘short-term’ impacts, collected at a local level did not systematically feed into NESS’s findings on the impacts of Sure Start. While NESS did report on local evaluation findings, for example through synthesis documents, this reporting was distinct from the national evaluation work to formally measure the impacts of Sure Start. A formal mechanism to ensure that such local ‘short-term’ impacts, that fed into NESS’s pool of data on the impacts and outcomes of Sure Start, would have been extremely valuable.

Another suggestion for facilitating greater integration of the national and local levels is that local programmes be encouraged and supported to robustly evaluate parallel themes or issues to those being investigated by national evaluation teams. If national evaluation teams worked to ensure that such ‘parallel’ local evaluation work was conducted using high-quality data-collection methods and tools, local evaluation data could more easily be incorporated into work of the national evaluation, while still meeting local needs.
The New Labour Government that implemented the Sure Start initiative recognized the importance of local communities and partners playing a central role in devising solutions to local problems, and flexibly re-shaping services in line with identified local need (Blair, 2000). New Labour was keen to devolve some power and decision making in the delivery of public services to a local level, and the Sure Start initiative was one example of this (Rouse 2001). At the same time, New Labour was also characterized by a target-driven focus and the implementation of centrally imposed monitoring and performance management frameworks (Flynn, 1999). The Sure Start initiative’s dual-level evaluation, and the lack of integration between levels, reflects this tension in New Labour’s promotion of local-level empowerment and simultaneous desire to centrally manage and control ‘performance’ and measure ‘outcomes’. In this context, it is perhaps easier to see why there was a disconnection between the local and national evaluation levels. SSU guidance suggests, initially at least, a ‘local’ focus for Sure Start local evaluations. Local evaluations were tasked with conducting work (e.g. evaluations of ‘process’) to enable the development and improvement of local services, but in keeping with their local focus, had little broader remit in terms of assisting in the evaluation of the Sure Start initiative as a whole. In contrast, the task of collecting the ‘evidence’ of Sure Start’s national impact was undertaken by a large, centrally organized national evaluation.

In future, it is important that the national and local levels of CCIs are better integrated, if the outcomes of these initiatives are to be effectively evidenced. This is particularly the case given the current Government’s drive towards ‘localism’, which aims to shift power in local services away from ‘central government’ and towards ‘local people’ (Department for Communities and Local Government, 2010). In this context, there is a danger that local evaluations of CCIs might become more ‘local’ in focus, and more disconnected from each other and from their national evaluations. Allen and Black (2006) also highlighted the need for greater integration between these local and national levels of Sure Start evaluation, drawing attention to the possible benefits of local evaluation teams’ involvement in collecting data on behalf of national evaluations. A recent article by Johnson and Weiss (2011) presents an example of how a multi-site initiative, with local and national levels of evaluation, can approach a greater degree of integration of evaluation levels. It is clearly possible for national evaluations of initiatives such as Sure Start to allow for local autonomy in evaluation, while also integrating local evaluations into national-level work to demonstrate impacts and outcomes. Bringing this about requires both a recognition of the disconnection that is common between local- and national-evaluation levels, and a top-down commitment to greater integration of local and national evaluation from those funding projects. Facilitating the effective coordination of the local and national levels is logistically difficult, particularly for a programme as large as Sure Start; however, the benefits, in terms of allowing for more effective impact and outcomes evaluation, would be potentially great.

**Increased co-ordination and ‘joining-up’ of local evaluations**

The predominantly individualistic nature of SSLP evaluations was a barrier to effective impact evaluation, and increased ‘joining up’ of SSLP local evaluations would have facilitated the larger-scale collection of data on the impacts of Sure Start. There was a need for greater facilitation and coordination of cross-programme evaluations, so that different SSLPs might evaluate similar aspects of their service delivery or similar types of impacts. This would have helped to facilitate the pooling of locally collected evaluation data, to provide a picture of Sure Start impacts across cities, regions, or larger geographical areas.
One way that this might have been achieved was to provide SSLPs with key impact-evaluation themes that all local evaluations should focus on, as well as effective guidance and tools for measuring the impact of services and activities. This would have helped to ensure that SSLPs were conducting evaluation work on similar themes, and if similar tools and methods were used across evaluations, there would have been greater scope for joined-up evaluation.

In addition, greater joint evaluation would have been facilitated, had all SSLPs been encouraged to conduct evaluation work across programme boundaries at an early stage in their development. Although all SSLPs were unique, in terms of their service delivery and populations, effective impact evaluation might have been encouraged by greater focus on the commonalities and common threads that linked programmes’ approaches to meeting the needs of their communities, and less focus on the differences between programmes. As stated earlier, the unique nature of each SSLP appears to have been a key factor in the SSU’s drive for ‘individualized’ local evaluations.

The move from SSLPs to Children’s Centres also provides important learning for encouraging joint impact evaluation. With the move from SSLPs to Children’s Centres, local authorities were tasked with delivering Children’s Centres within their authority area (DfES, 2006a, 2006b). Children’s Centres within each local authority area were therefore usually coordinated or managed by the local authority, and the fact that neighbouring Centres often operated under the umbrella of a single local authority led to an increase in the number of evaluations across Centres/Programmes. Joint evaluation meant that the scope of evaluations could be broadened. Local authorities were able to explore the impacts of their service delivery across larger geographical areas and populations, and make comparisons of the impacts of their work in different localities.

More recently, however, there has been a growth in local authorities commissioning out delivery of their Children’s Centre provision to external providers. In terms of evaluation, there is a danger that this commissioning out, particularly where different Centres within a local authority area are run by different organizations, may inhibit effective joint impact evaluation as Centres once more take more individualized approaches to evaluation.

Drivers and incentives for relevant partners to participate in evaluation

The difficulties in involving all relevant partners in the evaluation of a CCI are well-documented, and a key difficulty can be encouraging the involvement of all relevant stakeholder partners (Kings Fund, 2004).

Our work with Sure Start has further highlighted the importance of encouraging partner organizations to be involved in evaluation processes, if impacts and outcomes are to be effectively evidenced. Factors that we have highlighted as influencing partners’ willingness to actively participate in evaluation are similar to those that we have been found to influence willingness to participate in joint and collaborative working more generally. These factors include partners’ perceptions of the benefits to themselves or their roles, partners’ workloads, and partners’ understanding of what involvement in evaluation will mean for them (Harrington and Lloyd, 2008b; Lloyd and Harrington, 2008).

It is important then, that drivers and incentives exist to encourage relevant partners to engage in local evaluations. Where SSLPs were successful in encouraging partner involvement in evaluation, they took a number of approaches. For example, in some cases participation in evaluation was a requirement of funding and/or was written into SLAs. In other cases, partners were invited to workshops or training days on the topic of evaluation alongside programme staff, to encourage a shared understanding of evaluation and its importance. Whatever strategy is employed, it is important that partners actively engage in the evaluation process, and drivers and incentives to
participate can be important in bringing this about (Harrington and Lloyd, 2008b; Lloyd and Harrington, 2008).

**Improved local evaluation guidance**

SSLPs tended to look towards ‘official’ Sure Start Unit guidance, for information about what they should be evaluating and how. It was clear from discussions with many SSLP managers and local evaluators, during visits to programmes, that there was a need for clearer local-evaluation guidance as they were often unclear about what Sure Start Unit guidance required of them, in terms of evaluation. In some cases, this lack of clarity resulted in SSLP managers delaying evaluation work, or becoming disillusioned with the evaluation process per se. Key areas of uncertainty included whether to focus on evaluating ‘process’ or ‘impact’, when to begin evaluating services, what evaluation methodologies and methods to employ, and who should coordinate local-evaluation work.

Clear evaluation guidance is vital, since without this, those managing local evaluation from within CCIs can become wary of moving forward with evaluation work, for fear of proceeding incorrectly – this was certainly the case with some SSLPs. Delays in conducting evaluation work sometimes meant that programmes missed the opportunity to collect important baseline data, and thus the opportunity to use certain methodologies that would have yielded impact evidence. In our experience of Sure Start evaluation, clear evaluation guidance provides programme staff and evaluators with a level of certainty that helps to facilitate evaluation work.

Although it was important that individual SSLPs retained some flexibility to evaluate according to the requirements of their communities, greater prescription in terms of what areas of work to focus evaluate around and what tools and methods to use would have facilitated a greater amount of impact and outcomes evidence. A number of SSLP managers and staff who attended regional NESS Support evaluation workshops highlighted the need for clearer guidance about what they should be evaluating and how, and in some cases for greater Sure Start Unit prescription about evaluation.

SSU guidance (Sure Start, 2002) clearly stated that responsibility for answering questions about the impacts of Sure Start and what difference Sure Start made to the children who experienced it, lay with NESS rather than local programmes. The initial lack of any expectation that SSLPs would engage in outcomes evaluation that was particularly evident early in the roll-out of Sure Start, meant that local programmes were not provided with clear, written guidance about how to implement and embed systems and processes that might enable them to collect data on the impacts of services. Thus, impact evaluation was rarely well integrated into the service-planning process. Approaches such as ‘Theories of Change’ (ToC) (Connell and Kubisch, 1998) have the ability to help here by providing a framework and process through which those delivering services can gain an understanding of the links between activities and outcomes that can help them to implement more effective outcomes evaluation strategies, processes and practices (Lafferty and Maloney, 2003; Weiss, 1995). Such ‘theory-based’ approaches to evaluation were, however, rarely implemented in Sure Start local evaluations, despite their ability to help create an overall evaluation framework that can help stakeholders to engage with the evaluation process (Blamey et al., 2008).

**Improved local understanding of evaluation and greater local expertise**

Although the NESS Support module provided evaluation training, workshops and support for SSLPs and their partners, local understanding of the purpose of, and possibilities for, evaluation remained limited. There was a clear need to develop local understanding of what evaluation was,
how it could be used to garner evidence about impacts and outcomes, and what the potential benefits of effective evaluation were for staff, partners and local families. Without this, staff and partners often retained differing, and sometimes conflicting, perceptions of ‘evaluation’.

Developing a local understanding among the range of partners involved in SSLPs was a huge challenge. However, where SSLPs did attempt to foster a broad, shared understanding of evaluation, and develop evaluation skills and knowledge, for instance by providing evaluation training for a range of staff and key partners, evidence from discussions with programme managers and staff during visits to SSLPs and NESS workshops, suggest that there was often an increase in local evaluation activity.

The development of a shared, local understanding of evaluation, and the development of staff and partners’ evaluation skills, clearly had the potential to give rise to greater levels of local impact and outcomes evaluation. Certainly, those SSLPs where evaluators themselves had greater levels of evaluation skill and experience, and worked to develop staff and partners’ understanding of evaluation, tended to be more effective in evaluating the impacts of their services.

**Going ‘back to basics’ in developing evaluation systems and processes**

Those SSLPs that were most effective in collecting data on the impacts and outcomes of their services and activities tended to be those that took a systematic approach to evaluation, reviewing and modifying the systems, processes and tools that were in place within their programme, so that they could better capture evidence of programme impacts. These SSLPs attempted to implement an overall evaluation ‘system’ within their programme, rather than taking a more piecemeal approach to evaluation.

Such SSLPs were relatively rare, since effectively implementing such a programme-wide system could be relatively time-consuming and required the input and commitment of a range of staff and partners. Going ‘back to basics’ in this way, however, tended to prove more effective than attempting to evaluate impacts without the necessary infrastructure (e.g. in terms of monitoring data systems) being in place. Such an approach was particularly effective since it helped to embed good evaluation practice into on-going service-delivery and staff and partners’ on-going practice.

**Conclusion**

This article has highlighted a number of factors that inhibited the effective impact and outcome evaluation of the Sure Start initiative, and ways in which more effective evaluation of similar CCIs could be facilitated. In the case of Sure Start, the barriers to effective outcome evaluation highlighted in this article have meant that there is still relatively limited evidence about the impacts of the Sure Start initiative.

The example of Sure Start highlights the importance to effective outcome evaluation of those delivering national initiatives at a local level being appropriately supported in conducting their local evaluations, for example, through the provision of good quality evaluation guidance or appropriate evaluation expertise. It also highlights the need for appropriate drivers and incentives to encourage key partners to participate in the evaluation process.

A key barrier to effective impact evaluation was the disconnection between the local and national evaluation levels – an over-reliance on the centralized, national level evaluation to collect impact evidence, without the adequate incorporation and integration of local evaluation teams and evidence. The example of Sure Start suggests that greater integration of these evaluation levels may be one of the key routes through which impact evaluation of national CCIs could be improved.
For example, it suggests that in order to bring about more effective impact evaluation of similar CCIs in the future, there is a need for those commissioning, designing, managing and implementing national evaluations, to better understand and appreciate the potential importance and value of local evaluation evidence to a national understanding of the impacts of initiatives. This understanding and appreciation needs to be reflected in the national evaluation methodologies employed, so that there is greater integration of the national and local evaluation and a ‘de-centralization of national evaluation’.

National evaluations that fully integrate the local level also have the potential to help overcome some of the practical, methodological and cultural barriers to effective outcome evaluation outlined in this article. Active involvement and integration of local evaluations may, for example, result in better local understanding of the potential for measuring impacts and outcomes, greater local prioritization of outcomes evaluation, and better embedding of evaluation within delivery plans and on-going practice.

Clearly, local autonomy and the ability to structure local evaluations in line with local need are important to the ability of local initiatives to develop effective services. However, local autonomy in evaluation is not incompatible with greater integration of local and national levels of evaluation. The effective coordination and integration of the local and national levels may be logistically difficult for some initiatives, but the potential benefits in terms of allowing for more effective impact and outcomes evaluation are great.

**Funding**

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

**References**


**Nigel Lloyd** is Director and Principal Researcher at NLH Partnership Ltd. Please address correspondence to: NLH Partnership Ltd, No 1 Liverpool Street, London EC2M 7QD, UK. [email: nigel@nlhpartnership.co.uk]

**Louise Harrington** is Director and Principal Researcher at NLH Partnership Ltd. Please address correspondence to: NLH Partnership Ltd, No 1 Liverpool Street, London EC2M 7QD, UK. [email: louise@nlhpartnership.co.uk]