



DEFINING, ACHIEVING AND EVALUATING COLLABORATIVE OUTCOMES: A THEORY OF CHANGE APPROACH

Valeria Guarneros-Meza
Department of Politics and Public Policy
De Montfort University
Leicester, UK
Valeria.guarneros@dmu.ac.uk

James Downe
Centre for Public Policy
Cardiff University
Cardiff, UK
downej@cardiff.ac.uk

Steve Martin
Centre for Public Policy
Cardiff University
Cardiff, UK
martinsj@cardiff.ac.uk

Acknowledgements:

We are very grateful to the three anonymous referees, whose comments on earlier drafts helped improve this paper.

Abstract

Governments have repeatedly claimed that collaboration improves public service outcomes. However, defining, achieving and evaluating collaborative outcomes is often problematic. Analysis of multi-sectoral projects in Wales, which were supported by the European Social Fund, exemplifies these challenges. Shifts in policy discourses and the interplay between national and local agendas produced complex and contested understandings of outcomes which made difficult to evaluate the projects' achievements. We argue that the pursuit of collaboration needs to be understood not simply as an attempt to improve public service effectiveness but also 'cultural efficacy'. The conclusions offer reflections relevant for theory and practice.

INTRODUCTION

Despite the lack of clear evidence that collaboration leads to desired outcomes, governments continue to emphasize its importance in the design and delivery of public services (Dickinson and Sullivan, 2014; Denhardt and Aristigueta, 2011). The task of evaluating whether collaboration works is complicated by the complex and contested nature of performance in public service settings and the difficulty of drawing general conclusions from studies which focus on single policy areas and particular geographies of place.

This paper analyses barriers to defining, achieving and evaluating outcomes of projects that are characterized by multi-sectoral collaboration and inter-governmental working. The identification of barriers works as a mechanism to test whether national and local stakeholders' understandings of outcomes are grounded in evidence or reflect an untested assumption that collaboration improves the performance of public services. We argue that in order to identify barriers, it is important to recognize not only the efficiency and effectiveness aspects of performance, but also 'cultural efficacy' (e.g. policy discourse and practice) (Dickinson and Sullivan, 2014). In so doing the paper responds to the criticism that academics and practitioners often know more about the politics of a reform than about its effectiveness (Ashworth et al., 2010; Laegreid et al., 2014; Pollitt, 2009) because of the difficulty of defining outcomes during the process of evaluating performance.

The paper starts by briefly reviewing the literature on performance, the measurement of collaborative outcomes in public service settings and the use of theories of change in this area. Next we describe the European Social Fund-Local Service Board (ESF-LSB) Project and the Welsh policy context in which this was conceived and operated. The paper then describes the methods used to examine the Project's outcomes before presenting our findings which highlight the difficulties

that key actors experienced in defining outcomes and their perceptions about what collaboration achieved. We conclude by highlighting barriers (timing, scale and funding) to defining and measuring outcomes and the implications for researchers seeking to evaluate them.

DEFINING AND EVALUATING PERFORMANCE IN COLLABORATIVE ARRANGEMENTS

Performance in the public sector is a contentious concept because of the multiple and differing values adopted by diverse actors (Dickinson and Sullivan, 2014). Performance may include the quantity and quality of service outputs, consumer satisfaction, service objectives (targets and measures), expenditure data, equity in allocation of resources, and service outcomes (Andrews and Boyne, 2010). Interpretations of what constitutes a desirable public service outcome, and thus which of these dimensions of performance comes to the fore, are moulded by pragmatic, political and social factors which differ between contexts and among actors (Pollitt and Dan, 2013). In other words, performance involves power dynamics. 'As power alters between groups over time so do performance criteria' (Ashworth et al., 2010: 5).

The literature features three common models of assessing performance:

The goal attainment model suggests that performance should be judged in terms of the realization of objectives framed by policy interventions (final outcomes). This model is concerned with the extent to which service improvement leads to better outcomes for the citizen. However, not all goals of public interventions are easy to define. In many cases objectives remain ambiguous or abstract, and measurement is clouded by attribution problems and the combination of time lags and the limited timescales in which interventions are implemented and, consequently, evaluated (Heinrich, 2012).

The performance target model focuses on the creation of targets and indicators that measure issues such as quantity, quality, efficiency (cost-benefit) and equity of a specific intervention. Targets and indicators generally reflect outputs achieved over a determined period of time. This approach has been criticized for distorting behaviour, for example encouraging public servants to sacrifice quality in favour of narrowly defined short-term targets (Hood, 2006).

The processes and practices model focuses on the steps that pave the way to final outcomes. Depending on the level of resources invested to innovate or carry out specific processes, some academics consider that these can be classified as intermediate outcomes (Klijn et al., 2010; Edelenbos et al., 2013) which can include the adoption of best practices and innovative leadership, skills or organizational arrangements that help to correct procedures. The downside of this approach is that

because it does not focus on final outcomes it says little about improvements for the citizen. This leads to 'confusion between the antecedents of service improvement and the improvement of the service itself' (Ashworth et al., 2010: 4).

Dickinson and Sullivan (2014) offer a new perspective on performance to discourses dominated by efficiency and effectiveness. They argue that to obtain a more rounded and complete understanding of performance, cultural efficacy must also be taken into account. They emphasize factors such as rhetoric, emotions and symbols, which are materialized in actors' practices through their interpretations of meanings. In this paper, we argue that cultural efficacy is particularly relevant to the study of collaborative performance in Wales where, in common with many countries, collaboration has become a policy instrument of choice in spite of a lack of definitive evidence that it produces improvement for service users or citizens.

The literature on performance highlights the complexity of defining outcomes even within a single organization. Not surprisingly, the problem is accentuated in multi-agency projects associated to collaboration. In this paper, we understand collaboration as cross sector partnerships involving different tiers of government, together with business, non-profits or communities (Bryson, et al., 2006) in which collaborative management is an important factor in facilitating the partnership's multi-organizational operation (McGuire, 2006). In the complex micro processes that collaborative arrangements entail, conflict and tensions emerge as result of partners' different perceptions of and different organizational approaches to defining the problem, its solution and durability (Huxham, 2003; Klijn et al., 2010; Rittel and Webber, 1973). These challenges have come to the fore in the UK in the last two decades as successive governments have emphasised partnership working as a means of reducing organizational fragmentation in public service, tackling 'wicked issues', encouraging innovation and gaining access to new resources.

The literature on networks and collaboration, highlights ambiguities and uncertainties about what collaborative arrangements actually achieve (Laegreid et al., 2014). This poses a challenge not only for academics, but also for practitioners who aim to develop strategies and programmes that produce evidence-based results. Existing research addresses collaborative performance in two distinctive ways. The first focuses on management styles and their relationship with outputs and process-outcomes and shows how network management has a significant impact on *process* and *content* outcomes (Klijn, et al., 2010: 1066; Edelenbos, et al., 2013; McGuire, 2006).

A second approach shows how structure and context encompass resources and shape political opportunities and constraints. The seminal work of Meier and O'Toole (2001) examines the relationship between network management and outcomes (understood as service improvement in education through higher exam scores). They are also concerned with how management interacts with other factors that affect performance such as the external environment (climate and turbulence)

and previous legacies of hierarchic governmental arrangements. They contend that management cannot be isolated from resources (especially when there is financial dependency between actors) or institutional constraints (Meier and O'Toole, 2001). Similarly, Provan and Milward (1995) show how successful outcomes of a healthcare inter-organizational network depend on the centralized structure of the network (one single agency managing the network's daily operations), and a non-fragmented, stable and resource-rich environment provided by higher levels of government.

More recently, a meta-review of the literature on the effectiveness of collaboration by Turrini et al. (2010) brings the two approaches together. They acknowledge the relevance of contextual variables, alongside functional variables (behavioural characteristics which range from management strategies to beliefs held by managers) and structural variables (network organizational characteristics to achieve joint working and sustainability over time). Other authors reach similar conclusions and emphasize that service improvement depends on the external environment, organizational characteristics and organizational strategies to achieve outcomes, defined as service improvement (Cristofoli et al., 2015; Pollitt and Dan, 2013).

Dickinson and Sullivan's emphasis on cultural efficacy complements these analyses highlighting the importance of external environment and context. Dickinson (2014: 75) argues that in assessing partnership working, cultural efficacy will prompt evaluators to answer questions such as: 'What discourses of collaboration are present? How actors perform collaborative self? What are the affective dimensions of these discourses? And what type of symbols are present?' A good understanding of the political, economic and social context is required to answer these questions. In contrast, mainstream debates on efficiency and effectiveness tend to ask: 'What forms of partnership exist? Do partnerships lead to improved services and to improved outcomes for users? Are they cost effective?' (2014: 68)

Given the complexities of collaborative arrangements, 'theories of change' (ToC) have been considered a means of evaluating their outcomes. They help to assess collaboration because they acknowledge organizational synergies and coordination, the diverse assumptions held by stakeholders, and the production of knowledge and learning to overcome misalignment between partners' understandings (Connell and Kubisch, 1998; Healey, 2006). In so doing they open the 'black box' (Dickinson, 2008) situated between inputs and outcomes by asking 'how' and 'why' of the practices carried out throughout the process. In analyzing these practices, it may be found that causation of change is iterative and non-linear (Dickinson, 2008).

ToCs were originally employed in the United States to assess how change occurred in community initiatives and brought citizens and practitioners together to generate the theory. However, in the UK, they have commonly been used in government commissioned evaluations where 'expertise' is prioritized, rendering the

process more researcher-led (Mason and Barnes, 2007). In principle ToCs should be developed at the outset of an intervention, but in practice in the UK it is common to find that they are developed by evaluators commissioned after interventions are already under way (Downe et al., 2012). Although retrospection can offer advantages, there is a risk that these ToCs do not reflect all partners' perspectives (Marris and Rein, 1972; Mason and Barnes, 2007). Evaluators aim to bring together stakeholders' views to create a meta-narrative. But this process can risk overstating consensus by emphasizing what is politically acceptable and deploying measurement over deliberation as a better sign of credibility (Weiss, 1995).

The review presented in this section provides an analytical framework to analyse an evaluation of the ESF-LSB Project. First, we have valued debates on performance which have contributed to distinguishing among inputs, outputs and outcomes. Second, in acknowledging the debates that juxtapose collaboration and performance, we not only identify the importance of efficiency and effectiveness in collaborative arrangements, but also the importance of context. Third, in recognizing the importance of context, we highlight links to policy discourse and their enactment (cultural efficacy) as they feed into understandings of collaboration in Wales.

COLLABORATIVE PERFORMANCE IN WALES

Since 1999, responsibility for most public services in Wales (including health, local government and education) has been devolved to the Welsh Government. It has embraced a policy discourse which espouses the virtues of collaborative approaches to public service delivery and has been reflected in numerous policy statements and reports (e.g. Welsh Assembly Government, 2004; Martin and Webb, 2009). There is evidence that some progress has been made in developing collaborative approaches (Martin et al., 2013; Simpson, 2011; Sullivan et al., 2012; WAO, 2012). Over time there has been an increased emphasis on collaboration not just between public services, but also between public sector and third sector providers (Oldbell 3, 2015). However, performance has been patchy and organizations with their own distinctive missions, performance targets and lines of accountability, have found it difficult to work together (Martin et al., 2013).

In an attempt to accelerate the collaborative agenda, the 2011-2016 Welsh Programme for Government reiterated the importance of partnership working as a means to strengthen local democracy, achieve continuous improvement in public services, and develop more efficient and effective forms of service delivery (Welsh Government, 2011). More recently, there has been an increased emphasis on the importance of collaborative outcomes and the need for performance management of partnerships (Welsh Government, 2014a; 2014b). We discuss below how this has featured in government officials' practices through our evaluation of the ESF-LSB Project.

The ESF-LSB Project

The ESF-LSB Project was a product of this strong emphasis on collaboration. It aimed to improve the effectiveness and efficiency of public services by building up the capacity for collaborative working at local level among public and third sector organizations. Designed and managed by the Welsh Government, it was funded by the European Social Fund from which £17 million was made available to a series of local collaborations. As an EU-Priority 4 funded project, the Project had to demonstrate that service providers were supported by skilled managers, while the organizational capacity of such providers was being strengthened. The Project ran from January 2011 to December 2014 and included 38 'local delivery projects', which were collaborations between local authorities and other service providers aimed to improve service outcomes in a wide range of policy areas, such as ICT, health and social care, employment, transport, and housing.

Delivery of the Project's objectives was the responsibility of Local Service Boards (LSBs), which were set-up in 2008 as strategic partnerships of public service providers and third sector organizations to address 'wicked issues' which required coordinated and collaborative action. LSBs operated as a cross-service leadership teams to connect the whole network of public services in each local authority area (Welsh Assembly Government, 2007) and have recently been superseded by Public Service Boards (Welsh Government, 2015). This paper draws on an evaluation which tested whether the Project delivered improvements in public services and outcomes for citizens. Efficiency and effectiveness of partnership working dominated the evaluation; partly because it was requested by the research funders and also as a result of the need for evidence to assess whether collaboration can produce service improvement. However, the ToC provided an opportunity to 'look behind the scenes' and analyse the 'hows' and 'whys' of practices that feed into understandings of partnership; enabling us to tap into cultural efficacy and consider its salience.

DATA AND METHODS

To identify the intended outcomes of the ESF-LSB Project we developed a ToC through in-depth discussions with senior policy-makers, responsible for managing the Project, and local stakeholders (Figure 1). The theory's development took place a few months after the Project started and followed the 'expert' researcher-led approach mentioned above. Although this theory reflected the dominance of effectiveness, outcomes and service improvement found in the policy discourse, our interactions with government officials and local stakeholders (public and third sector officers and politicians and local evaluators) provided room for these terms to be questioned throughout the process.

[Figure 1 here]

The ToC incorporated the key concepts of performance of networks and partnerships identified in the literature. It was dynamic and could be revised as we tested its applicability through interviews with national and local stakeholders in seven local delivery projects. These projects were selected based on a large number of criteria including the amount of grant received; the number of local authorities and LSB's involved; coverage of key policy areas; the inclusion of the third sector in the provision of services; the geographical location; and perception of the political salience of the project. This paper analyses two – the North Wales-ICT Collaboration (NW-ICT) and Caerphilly Passport which reflected either of the ends of the spectrum in terms of understanding and achievement of collaborative outcomes and their alignment to the broader policy discourse. The other five delivery projects fell somehow between these two extremes.

The 'context' in which the Project operated encompassed four external (or environmental) factors which the literature highlights as drivers of collaborative outcomes. Resources, such as staff and finance, are important for ensuring that the discourse of outcomes reaches not only the national, but also the local understanding of partnership working. Monitoring is relevant when partnership working includes the participation of national government as a partner (Martin and Guarneros-Meza, 2013). The stability of a collaborative discourse, since devolution, indicates its sustainability over time.

'Inputs' included existing partnerships (LSBs) and the financial and managerial resources of the Welsh Government which instigated the Project. We defined 'outputs' as steps on the way to building collaborative arrangements, which were the preferred mechanisms to achieve final outcomes in service provision. We also identified and distinguished between two types of outcomes: 'process' outcomes, which we defined as the evidence that indicates some degree of service improvement achieved collaboratively (e.g. cost savings or innovative arrangements), and 'citizen-oriented' outcomes, which were defined as the extent to which process outcomes lead to improvements for the citizen.

[Table 1 here]

We undertook documentary reviews of government reports and all local project reports, minutes of partnership meetings and logbooks. In addition, we collected data through 39 in-depth semi-structured interviews with key local and national actors. Between July and September 2013 we interviewed 15 members of the Project's national Advisory Board, including civil servants from the Welsh Government's health, education, social services and local government ministries, the health service and police, representative bodies of local government and the third sector, and the Big Lottery Fund. We then interviewed 19 local partners who were involved in the design and implementation of the two projects (10 in NW-ICT and nine in Caerphilly Passport). These interviews took place between December 2013 and June 2014. Five follow-up interviews were conducted in October 2014 to

provide an opportunity for national stakeholders from different organizations to reflect upon the outcomes generated by the Project.

Interview topic guides followed the ToC. We examined local actors' experience of implementing their project. Interviews with national actors examined the relevance of the Project to their organizations' objectives and their perceptions on what the Project was expected to achieve. Interviews were audio-recorded and transcribed. Coding was thematic and sought phrases and keywords which revealed interviewees' understanding of outputs and outcomes through accounts of instances of disagreement, compliance, mediation, resignation and learning.

The next section unpacks the evidence from our data about barriers to defining and achieving outcomes while responding to the ToC meta-narrative. The discussion begins with national stakeholders' perceptions of the problems of defining outcomes which were reflected in the case studies' outcomes and in turn fed back to Welsh Governments officials' understandings of the limitations that the ESF-LSB Project had in defining and achieving outputs. The quotes given are illustrative of the 'hows' and 'whys' that built or rejected common understandings.

DEFINING OUTCOMES AT NATIONAL LEVEL

Analysis of the data from interviews with national officials highlighted four main instances where problems in building an understanding about outcomes took place. The first was the increasing importance given to outcomes in the policy discourse as the Project was developed and implemented. A Welsh Government official familiar with the daily management of the Project explained:

The Welsh Government was changing and it was moving from the Efficiency and Innovation Board through to the Public Service Reform Agenda and starting to focus on what eventually became the Programme for Government. So we were then able to focus project development around 'is this going to help meet...[the] Welsh Government's aims and objectives and outcomes in the Programme for Government?' Would it link up with work we were already doing and support stuff around effective services for vulnerable groups? Would it fit in with the wider public service reform agenda?

The Project's Advisory Board, led by Welsh Government officials, had to adjust the selection process of the local delivery projects to reflect this. Over time they placed greater emphasis on projects which aimed to have direct impacts on citizens, and this was reflected in the allocation of funding for projects in 2012 and 2013.

The second source of tensions was the co-existence within the policy discourse of innovation through collaboration (in delivering services) and the need to

demonstrate improvements for the citizens. A Welsh Government official in charge of monitoring all local delivery projects suggested:

Back early in 2010, [name] was in charge of the Project back then... She came to talk to me for a prospect of doing some sort of evaluation. She had in mind a small process evaluation just sufficient to satisfy the [Welsh European Funding Office's] requirements which was - any project of over £2m needs an external evaluation of some sort. But when I read the business plan it was a lot more complex than a process evaluation and would have been a missed opportunity if we didn't look at outcomes, improved service delivery. I prepared a scoping paper and they were a bit surprised about the level of evaluation I was proposing. It didn't get resolved and everything went quiet for about six months until [new Project manager] came on board...He talked to me and was very receptive to my ideas.

The conflation between collaborative innovation and citizen outcomes resulted in indecision about how much of the Project's funding was going to be dedicated to evaluation as well as on the requirement for local projects to dedicate part of their funding to local evaluation (external or in-house) to ensure that outcomes were achieved. The receptiveness of the new Welsh Government Project manager to the importance of evaluation meant that more funding (1.5% of the total budget) was dedicated to this purpose. However, this was not enough to cover all costs and training of local stakeholders to manage or carry out their own evaluations. In effect, the requirement that all local delivery projects had to undertake their own local evaluation was not clearly established until early 2012, approximately a year after some had begun.

The delay in establishing the importance of outcomes and their evaluation led to local project managers setting objectives which were often not sufficiently specified. This third problem was exacerbated by the Welsh European Funding Office (WEFO) indicators which, rather than setting outcomes, measured activities and outputs such as the number of dissemination initiatives; collaborative agreements among service providers; secondment placements made available; and organizational learning strategies produced. This created difficulties for Welsh Government officials managing the Project and for local stakeholders who initially thought it was enough to report outputs. It also wasn't clear if and how the WEFO indicators related or contributed to the intended outcomes.

The fourth set of problems which were identified from the interview data related to the lack of transparency among members of the Advisory Board about the criteria used to assess if proposed delivery projects were feasible and robust enough to achieve outcomes and receive funding. One former Board member with a local government portfolio explained:

So I think in terms of quality I found it really, really difficult to make any judgement. I could take bits out of an application and say, 'Outcomes not very clear'... but to be perfectly honest...only [name] from [agency] had the

skills to sit on that panel and make judgements about what were good projects and what were bad ones...My view is that the civil servants who sat on that panel were too quick to reject certain projects that maybe just collided a little with- They were maybe being funded elsewhere or the objective was part of some policy in their department that only they knew the minutiae of- and therefore we had to take them at face-value that there was some kind of conflict there.

This view was supported by a government official who explained that 'project bids were scored by Welsh Government officials from silo-ed teams'. This implied that although collaborative outcomes were being asked of local stakeholders, government officials were not doing enough to work collaboratively at national level and instead were funding projects which supported their own departments' priorities.

In the early stages of the Project, we found that members of the Advisory Board expected local projects to develop innovative collaborative arrangements which had potential to be rolled out more widely through the dissemination of best practice, but were only expecting a handful of projects to achieve improved outcomes for citizens. Meanwhile, they continued to conflate innovation and outcomes. For example, a government official in charge of performance believed that as the local projects 'are not very innovative...the expectation of improved outcomes is unlikely'. This disappointment, which permeated other members of the Advisory Board, led some to lower their expectations with regard to the size, scale and significance of local delivery projects. As one put it:

If a project makes an impact to 50 vulnerable people with complex needs, then that is a success. If a project provides a [service delivery] 'model', then that is also helpful. Not very high expectations for millions of pounds of investment!

The four tensions revealed by the interview data highlight symbolic behaviours that recreate the value of efficiency and effectiveness of traditional performance. The preference for outcomes after periods of silence and ambiguity, the promotion of WEFO targets which didn't measure outcomes, the allocation of insufficient resources for local evaluations and the making of decisions favoring outcomes despite the lack of skills of some people in the Advisory Board show the importance of the performance discourse, albeit poorly conceived.

LOCAL STAKEHOLDERS' UNDERSTANDING OF OUTCOMES

In this section, we analyse how outcomes were understood by local stakeholders in the two delivery projects. The NW-ICT project was designed to deliver cost efficiencies through the technological modernization of ICT services across six local authorities in North Wales. The Caerphilly Passport project aimed to reduce the number of young people who were not in education, employment or training (NEET) through internships and mentoring before being 'passported' into employment

opportunities. Table 1 provides a comparative summary of these projects. In both cases, approximately three-quarters of the funding was devoted to the salaries of the project manager (and support staff) responsible for its implementation. This reflects the importance placed on building collaborative capacity through management as a means to develop better quality services.

The NW-ICT project was designed and funded in the first tranche of projects in 2011 when there was less emphasis on outcomes in the policy discourse. The Caerphilly Passport project was commissioned in 2012-2014 by which time the importance of outcomes for citizens was better defined and understood by Welsh Government officials. The difference in timescale not only affected the design of each project and its approach to evaluation, but also the level of support local stakeholders had towards the design of our ToC. In effect, the greater clarity that Welsh Government officials had on process and citizen-oriented outcomes, the greater the 'buy in' to the ToC by local stakeholders. The regional coordinator overseeing NW-ICT responded to our questions regarding outcomes on service improvement and citizen impact in a frustrated and ironic way:

We delivered the project as set by its objectives [to deliver cost efficiencies through technological modernization of ICT services in North Wales] and you are coming as evaluators with a...set of tools that have different criteria [impacts on the citizen] that were never discussed in the beginning and it is bound to be a mess, isn't it?

In contrast, the Caerphilly Passport project clearly aimed to make a difference to citizens and commissioned its own external evaluation which developed a ToC approach with outcomes split between those likely to be achieved in the short, medium and long-term.

Collaborative outputs

All local delivery projects were required by WEFO to set targets in four main areas - number of dissemination initiatives; collaborative agreements; secondment placements; and organizational learning strategies. These targets were achieved in the NW-ICT project which delivered six collaborative agreements against the target of one. The project manager was clear that these outputs were the vehicle that consolidated collaboration among a myriad of public and third sector organizations. Most of the manager's time was spent lobbying partners and liaising with them to reach agreements. The project manager was able to convince all parties to reach agreements, sign business plans, and distribute responsibility among partners to begin innovative ways of working together.

The performance of the Caerphilly Passport project on the WEFO indicators was more mixed. It was expected to achieve six collaborative agreements but delivered only one. The project exceeded targets relating to secondment positions and dissemination initiatives. The WEFO indicators were described, however, as

being 'pointless – it's ticking the box'. Respondents were unsure about what these targets were intended to achieve and they were not used to monitor the progress of the programme. Instead, the focus of the project was on achieving targets that the Passport team had set themselves based on previous work conducted by the council and its partners in this policy area. These targets were designed to be 'realistic and deliverable' given the context and the partnership that they had to build on. Table 2 shows how the project delivered on these outputs.

[Table 2 here]

Collaborative process outcomes

The NW-ICT project's main process outcomes were cost savings made by its key partners (six local councils). Different work-streams in the project achieved total savings of approximately £262,000. Despite this success, interviewees suggested that savings were not maximized because budgetary restrictions imposed by local politicians were not aligned across the six councils. Instead, cost savings were achieved partially across the different work-streams of the project when budgets and organizational arrangements of two or three local councils had similar structures and did not have to undergo substantial change.

Cost savings were not an aim of the Caerphilly Passport project, however, it can be seen as representing value for money given that the average individual life-time cost to the public exchequer of a NEET person is estimated to be £56,300 (Coles et al. 2010). The project led to improved awareness of partners' capacity and opportunities for collaboration. Partners, including local businesses, became more familiar with each other and this led to process improvements such as the introduction of multi-agency working groups and collaborative funding and resource between uncommon partners such as the local council and JobCentre Plus (a 'quango' helping those attempting to find employment).

Citizen-oriented outcomes

Interviewees from the NW-ICT project were clear that modernization of ICT through cost savings was an 'enabler' to develop outcomes that could have an impact on citizens further down the line. For example, they hoped that technologies would facilitate new forms of service delivery leading to improved citizen satisfaction. This was particularly evident in a work-stream that aimed to pilot an online database used by social workers belonging to the Emergency Duty Team (EDT). The piloting exercise was seen as successful because the improvement in IT arrangements reduced the time needed to access information by social workers; facilitated access to the database remotely; and extended the access to the database by all frontline staff.

But the lack of emphasis on citizen outcomes meant that nobody monitored the benefits of this change for service users. The ICT team suggested that it was

beyond their remit and that the EDT should be responsible. But the EDT manager, alluding to the downsides of a target culture, stated, '[monitoring] will take a lot of work and we will not see a benefit from it in terms of service'. Finally, the NW-ICT project manager stated that the monitoring of these issues was beyond the project's remit and there needed to be a steer from the LSB. While LSB members were aware of this new service change, they were busy designing strategies to cope with expenditure cuts resulting from the 2010 fiscal crisis.

The Caerphilly Passport project aimed to provide a cohesive, holistic support programme for young people (aged 16-24) which could potentially have an impact on several outcomes including assisting LSB partners to address workforce planning issues, ensuring local skills gaps are addressed, and ultimately providing employment opportunities for young people. By the end of the project, 177 participants completed the Jobs Growth Wales[1] work placement and 80% of participants had a positive outcome (137 went into full time employment and four into education).

The support provided by the Passport model, through several training opportunities, helped young people to become 'work ready'. They were able to develop existing skills and abilities which led to practical outcomes such as improved interview skills or an enhanced CV. In addition, surveys of participants revealed improved social skills, increased levels of self-confidence and motivation to enter the job market. Just over seven in ten participants reported that without Passport they would still be unemployed (Wavehill 2014: 79). The Passport model faced challenges as key stakeholders, such as the Health Service, were reluctant to change their organizational structures to fit into the project. Although the Health Service had 'come to the table on a regular basis', the project only focused on the local authority's administrative boundary, whereas health's covered the Gwent region, which encompasses four other local authorities.

THE REACTION OF WELSH GOVERNMENT OFFICIALS TO THE OUTCOMES ACHIEVED

The second round of interviews conducted with senior Welsh Government officials in 2014 enabled reflection on the extent to which delivery projects had achieved their intended outcomes. The consensus was that some outcomes for citizens had been achieved but only in a minority of projects, especially those that aimed at improving outcomes for children and families. Interviewees believed that projects aimed at achieving longer-term, system-wide change had been less successful.

The process outcomes achieved by the NW-ICT project in terms of cost savings fulfilled the Welsh Government's expectations. However, officials were disappointed that the project's stakeholders and LSB leadership did not incorporate in their design the impact that the project could have had on the citizen through for

example the EDT. The Welsh Government did not press for this because of its own initial indecision about the importance of outcomes and consequently the lack of definition of citizen outcomes at the time when the NW-ICT project was designed and implemented. This decision reinforced the belief that back-office functions can't have a direct impact on the citizen.

Welsh Government officials regarded the outputs and outcomes achieved by Caerphilly Passport as satisfactory. This can be evidenced in three ways. First, the Welsh Government provided flexibility in the way they allocated additional Jobs Growth Wales places to the council as opposed to a work-based learning provider. This was the first time that it had done this and provided a clear demonstration of national support for the council's approach. Secondly, the project was used as a case study in the Welsh Government's Youth Engagement and Progression framework. The report states that 'We are keen to extend this type of strategic approach more widely across the public sector in Wales' (2013: 51). Finally, the well embedded understanding of outcomes by the Passport project, fed back to national level on the irrelevance of WEFO indicators. Although these had to carry on being reported, the national team acknowledged their poor design.

By mid-2013, members of the Advisory Board started to question the extent to which local delivery projects would be sustainable after the ESF funding ended. In the case of NW-ICT, the project was dissolved after the funding stopped. There were some incipient signs that a few partner organizations wanted to assess how changes in ICT could benefit the citizen, but this was to be done on a single organizational basis as opposed to collaboratively. In Caerphilly, the Passport team secured further grant funding from the national Job Centre Flexible Support Fund that sustained the programme until March 2015, when it formally ended.

Our analysis of the contrasting perceptions and fortunes of these two projects shows the importance of cultural efficacy as reflected in alignment/non-alignment by delivery projects with the Welsh Government's expectations of them. The NW-ICT project set out to achieve cost savings through collaboration, as required at the time when it was designed, but local actors did not embrace the Welsh Government's increasing emphasis on the importance of demonstrating the achievement of citizen outcomes. This meant that it was only partially consistent with the ToC articulated by national actors, but the WEFO indicators, which did not measure outcomes, worked well in maintaining a collaborative discourse.

In contrast, the timing of the funding of the Caerphilly Passport project contributed to its full alignment to the ToC. It designed its own indicators to monitor and achieve citizen outcomes. The clearer definition of outcomes also led to more regular meetings between Welsh Government staff, the local project team and a local external evaluator. This built a common understanding of the ToC. As a result, the project was seen as a model for subsequent youth policies.

DISCUSSION

Over time, national actors increasingly emphasized that the ultimate aim of local delivery projects funded by the ESF-LSB Project was to improve service provision in order to produce better outcomes for citizens. However, in the early stages of the Project there was far less clarity about the meaning of outcomes and collaboration was seen as an end in itself. This reflects the collaborative discourse in Wales which has been present for more than ten years but has increasingly emphasized the importance of outcomes. The evaluation of the ESF-LSB Project, therefore required a ToC that acknowledged the importance of outputs, process outcomes and citizen-oriented outcomes.

The ToC became a means of exerting pressure on local projects to show how they would achieve outcomes. Government officials gradually began to meet more regularly with local stakeholders to ensure that outcomes were factored in to the design and implementation of local projects and, as in the Caerphilly Passport project, projects began to feed back to their national counterparts by perfecting the overall definition and design of the evaluation. The two projects we studied illustrated that there were three main barriers to defining, achieving and evaluating collaborative outcomes - timing, scale and funding - which are relevant to both theory and practice of collaboration in inter-governmental relations.

First, the initial indecision by Welsh Government officials to emphasize outcomes in the design of the ESF-LSB Project was costly because the first wave of projects did not recognize the need to achieve citizen-oriented outcomes. Thus, stakeholders in the NW-ICT project believed it was sufficient for them to achieve cost savings through collaboration. The project had an opportunity to make an indirect impact on citizens (i.e. through the EDT), but did not emphasize this. The lack of clarity about outcomes among government officials orchestrating the policy discourse meant that monitoring changes promoted by collaboration became nobody's responsibility. In contrast, once government officials had a clear focus on outcomes, the effectiveness dimension of performance emerged strongly (as in Caerphilly Passport).

Second, outputs (collaboration agreements) and process outcomes (cost savings and training sessions for young people) can be achieved collaboratively, but when outcomes have to be citizen-oriented and sustainable, there is a tendency to return to single-organization mechanisms as a way of achieving them, jeopardizing collaborative arrangements that span across local and regional scales. In the NW-ICT project, as each local council wanted to maintain its political and organizational autonomy, citizen outcomes were pursued independently by each of them according to local priorities and budget. In Caerphilly Passport, the lack of coterminosity between local government and health services threatened the durability of the

project because key actors were reluctant to change their regional organizational structures.

Third, the three-year funding meant that some delivery projects were focused on looking for continued funding, as opposed to changing systems for improving citizen outcomes. This raises serious questions on the nature of the ESF-LSB Project and whether the local delivery projects had been properly scoped to identify ambitious but realistic outcomes within this timeframe.

CONCLUSIONS

The ToC we deployed was co-designed with national and local stakeholders, alongside some local project evaluators. It helped to define outcomes and encouraged some projects to look beyond outputs to demonstrate that they had achieved collaborative outcomes which would benefit citizens. The process of developing the ToC led to more frequent communication between government officials and local project managers and indicators and realistic targets being set as steps towards measuring outcomes. The ToC came with a cost though, as some local actors believed that they were performing well only to find that national policy makers had a different set of expectations or, in their view, had 'moved the goal posts'.

At the end of the evaluation, collaborative performance was aligned to the discourse of effectiveness and efficiency. This resulted from the broader policy discourse that conflated collaboration and service improvement. This discourse was not only found in government reports, but also in the Project research commissioning documents and in the evaluators' familiarity with traditional debates on performance. The ToC reflected a researcher-led focus that top-down government initiatives in the UK commonly pursue. Although these are limitations of our approach, we also argue that the latter derive more generally from the characteristics of ToCs.

ToCs recognize a multiplicity of stakeholders' views and understandings, they are also designed to reach consensus over time through learning. This poses a threat because dominant narratives are likely to override other views at the end of the process. As a result, although ToCs are designed to be inclusive they do not guarantee equality of influence over the performance criteria that are adopted (see Healey, 2006 for minimizing this problem). In the NW-ICT project, discussion and learning between local stakeholders and Welsh Government officials broke down as differences in views about the type of outcomes to be achieved surfaced. In contrast, the shared understanding of outcomes between the Caerphilly Passport team and the Welsh Government officials strengthened support for the project and helped secure continued funding.

Our research supports claims that researcher-led ToC can misrepresent or downplay some actors' aspirations (Mason and Barnes, 2007; Dickinson 2008). Although the evaluation of the ESF-LSB Project was led by a national Advisory Board from a wide range of organizations, the ToC employed, became a means of shaping and articulating expectations of the delivery projects, which did not necessarily reflect local actors' objectives. The ToC was useful at defining collaborative outcomes, but rather than encouraging collaboration between government and local agencies, it reinforced a hierarchical relationship between national government and local partnerships.

Understanding the reasons for this highlights the value of cultural efficacy and the need to heed the discourse, practices and symbols that differing interpretations of effectiveness and efficiency portray. Cultural efficacy explains the frenzy for measurement that was created by the Project and our own position as evaluators in reinforcing discourses of effectiveness and efficiency in collaboration. We conclude that ToCs can play a valuable role in assisting policy makers and practitioners to think about and articulate how collaboration is expected to improve performance and to test whether it does so. They can also help to identify potential links between different dimensions of 'performance' (including process of collaboration, service improvement and citizen outcomes). But they do not necessarily overcome asymmetries in the power relations among actors and are liable to change over time in response to broader environmental or contextual specificities.

The use of a ToC is not 'neutral'. Measures of citizen outcomes remain contested because of the multiple understandings of performance that the intended outcomes entail. Hence a ToC may expose divisions in ways which disrupt collaboration. There is a need for research to better understand the leadership - individual, collective or shared across different sectors (government, academia and civil society) – and power relations among partners at national and local levels, and to identify strategies for the management of conflict during the definition, achievement and evaluation of collaborative outcomes.

Our research shows that multi-agency collaboration can achieve citizen outcomes if the need for this is clearly stated from the outset and the timing, scale and funding of collaborative projects is aligned to it. A ToC can act as a useful guide by helping to focus attention on outcomes that benefit citizens as well outputs and processes of collaboration, and by articulating and testing the links between them. This paper contributes to understanding of the difficulties that practitioners and scholars face in measuring the citizen outcomes of collaborative endeavours. It confirms the importance of an integrated approach to understanding the outcomes of collaboration (Turrini et al., 2010), which takes account of context, behaviour and structure, as well as for cultural efficacy (Dickinson and Sullivan, 2014) where the performance of values and beliefs reveal the rationale of the assessment public service.

NOTES

[1] Jobs Growth Wales began in December 2012 and provides unemployed young people with a job for six months with the intention that all jobs will be sustained by the host employer after completion.

REFERENCES

Andrews, R. and G. Boyne (2010) Better Public Services: The moral purpose of public management research? Discussants: Christopher Hood, Owen Hughes and Don C. Kettl, *Public Management Review*, 12:3, 307-321

Ashworth, R., G. Boyne and T. Entwistle (2010) An Introduction in R. Ashworth et al., (eds.) *Public Service Improvement: Theories and evidence*, Oxford: Oxford University Press, pp. 1-14.

Cristofoli, D., L. Maccio and L. Pedrazzi (2015) Structure, mechanisms, and managers in successful networks, *Public Management Review*, 17(4), pp. 489-515.

Coles, B. Godfrey, C. Keung, A. Parrott, S and Bradshaw, J. (2010) *Estimating the life-time cost of NEET: 16-18 year olds not in Education, Employment or Training, Research Undertaken for the Audit Commission*, York: University of York.

Connell, J. and A. Kubisch (1998) Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects and Problems, in K. Fulbright-Anderson, A. Kubisch and J. Connell (eds) *New Approaches to Evaluating Community Initiatives*, vol. 2, Theory, Measurement, and Analysis. Washington, DC: Aspen Institute.

Denhardt, K.G. and M.P. Aristigueta (2011) Performance management systems: Providing accountability and challenging collaboration, in W. Van Dooren and S. Van de Walle (eds.) *Performance information in the public sector*, Basingstoke: Palgrave-Macmillan, pp. 110-126.

Dickinson, H. (2008) *Evaluating outcomes in health and social care*, Bristol: Policy Press.

Dickinson, H. (2014) *Performance governance: Partnerships, culture and New Labour*, Basingstoke: Palgrave Macmillan.

Dickinson, H. and H. Sullivan (2014) Towards a general theory of collaborative performance: the importance of efficacy and agency, *Public Administration*, 92(1), pp.161-177.

Downe, J. Martin, S. and Bovaird, T. (2012) Learning from Complex Policy Evaluations, *Policy & Politics*, 40(4), pp. 505-523.

Edelenbos, J., A. Van Buuren and E.H. Klijn (2013) Connective Capacities of Network Managers, *Public Management Review*, 15(1), pp.131-159.

Healey, P. (2006) *Collaborative planning: Shaping places in fragmented societies*, Basingstoke: Palgrave Macmillan.

Klijn, E.H., B. Steijn, J. Edelenbos (2010) The impact of network management on outcomes in governance networks, *Public Administration*, 88(4), pp1063-1082.

Heirich, C. (2012) Measuring public sector performance, in B.G. Peters and J. Pierre (eds.) *The Sage Handbook of Public Administration* (2nd edition, London: Sage), pp. 32-49.

Hood, C. (2006) Gaming in targetworld: the targets approach to managing British public services, *Public Administration Review*, 66(4), pp.515-521.

Laegreid, P., K. Sarapuu, L.H. Rykkja and T. Randma-Liiv (2014) *Organizing for Coordination in the Public Sector: Practices and lessons of 12 European countries*, Basingstoke: Palgrave-Macmillan.

Marris, P. and M. Rein (1972) *Dilemmas of social reform*, second edition, Harmondsworth: Penguin Books.

Martin, S.J. and Webb, A. (2009) 'Citizen-centred public services: contestability without consumerism', *Public Money and Management* 29(2), pp. 123-130.

Martin, S. J. and V. Guarneros-Meza (2013) Governing Local Partnerships: Does External Steering Help Local Agencies Address Wicked Problems? *Policy & Politics*, 41(4), pp. 358-603.

Martin, S.J., J. Downe, T. Entwistle and V. Guarneros-Meza (2013) *Learning to Improve: an independent assessment of Welsh Government's policy for local government – final report*, Cardiff: Welsh Government.

Mason, P. and M. Barnes (2007) Constructing theories of change: methods and sources, *Evaluation*, 13(2), pp. 151-170.

Meier, K.J. and L. O'Toole (2001) Managerial strategies and behavior in networks: A model with evidence from US public education, *Journal of Public Administration Research and Theory*, 11(3), pp. 271-293.

Pollitt, C. (2009) Structural change and public service performance, *Public Money and Management*, 29(5), pp. 285-291.

Pollitt, C. and S. Dan (2013) Searching for impacts in performance-oriented management reforms: A review of the European literature, *Public Performance and Management Reviews*, 37(1), pp. 7-32.

Provan, K.G. and H.B. Milward (1995) A preliminary theory of interorganizational network effectiveness: A comparative study of four community mental health systems, *Administrative Science Quarterly*, 40, pp.1-33.

Rittel, H.W.J. and M.M. Webber (1973) Dilemmas in a general theory of planning, *Policy Sciences*, pp. 155-169.

Simpson, J. (2011) *Local, Regional, National: What services are best delivered where? A Report to Carl Sargeant AM, Minister for Social Justice and Local Government*, Cardiff: Welsh Government.

Sullivan, H. and C. Skelcher (2002) *Working across boundaries: Collaboration in public services*, Basingstoke: Palgrave.

Sullivan, H., P. Williams, S. Jeffares (2012) Leadership for collaboration, *Public Management Review*, 14(1), pp. 41-66.

Turrini, A. Cristofoli, D., F. Frosini, and G. Nasi (2010) Networking literature about determinants of network effectiveness, *Public Administration*, 88(2), pp.528-550.

WAO (2012) *Public Engagement in Local Government*, Cardiff: Wales Audit Office.

Wavehill (2014) *Final Evaluation of the PASSPORT Programme*

Weiss, C.H. (1995) Nothing as practical as good theory in Connell, J.P., A.C. Kubisch, L.B. Schorr and C.H. Weiss (eds.) (1995) *New approaches to evaluating community initiatives: Concepts, methods and contexts*, Washington, D.C.: Aspen Institute, pp. 65-92.

Welsh Assembly Government (2004) *Making the Connections*, Cardiff: WAG.

Welsh Government (2011) *Programme for Government*, Cardiff: Welsh Government

Welsh Government (2013) *Youth engagement and progression framework Implementation plan*, Cardiff: Welsh Government.

Welsh Government (2014a) *Commission on Public Service Governance and Delivery*, Cardiff: Welsh Government.

Welsh Government (2014b) *Devolution, Democracy and Delivery: Improving public services for people in Wales*, Cardiff: Welsh Government.

Welsh Government (2015) *Well-being of Future Generations (Wales) Act 2015*, Cardiff: Welsh Government.

Williams, G. (2015) *Independent evaluation of WCVA's Making the Connections Project*, Cardiff: Welsh Government.

Figure 1. Theory of Change

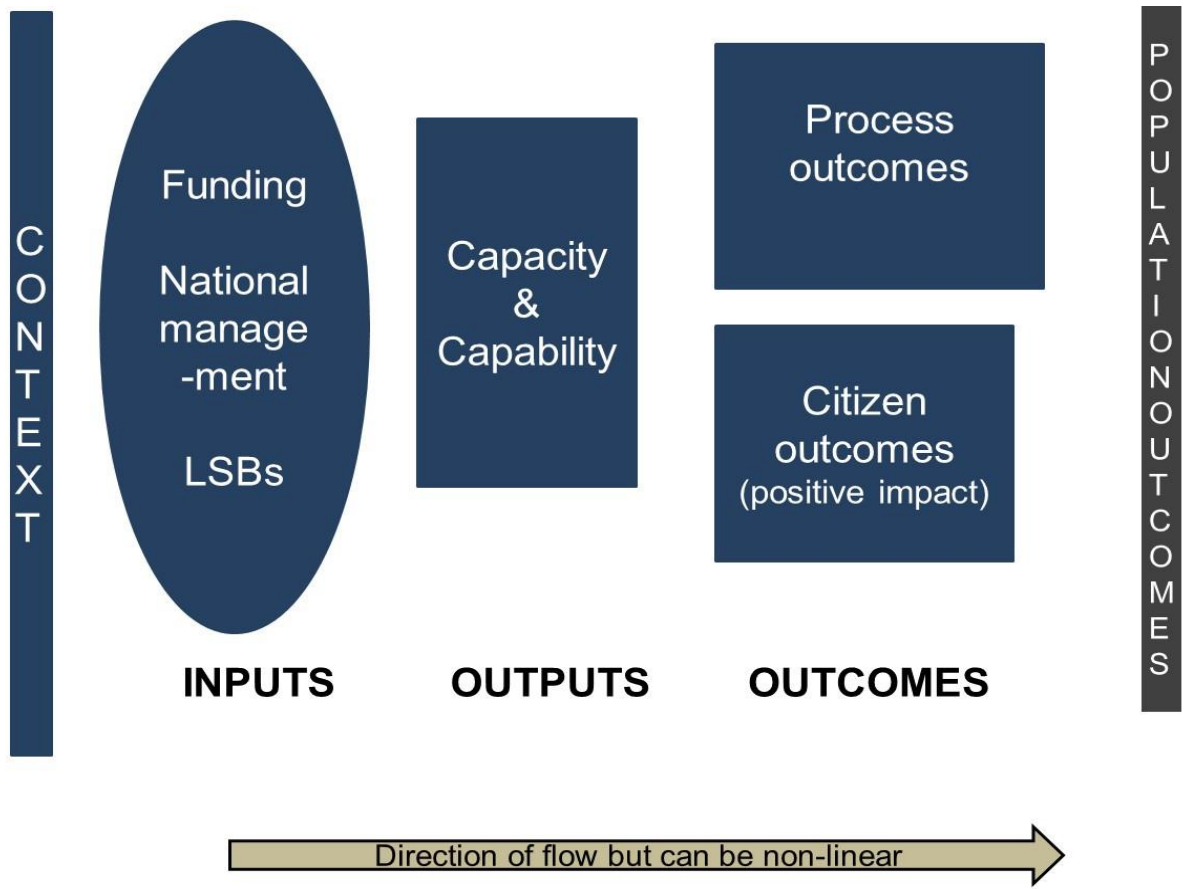


Table 1. Summary of ESF-LSB Project and the two case studies

	ESF-LSB	North Wales ICT (6 local authorities)	Caerphilly Passport (1 local authority)
Start-end date	January 2011-December 2014	April 2011-November 2013	November 2012-December 2014
Objectives	Achieve better public services and outcomes for citizens	Cost savings through ICT modernisation	Reduce the number of young people who are NEET
Partners collaborating	-----	11 partners including local councils, fire and rescue, health, police, regional university	6 partners including local council, working-age and life-long career support quangos, health, work-based learning provider, local business forum
Rationale of collaborative performance (theory of change)	Theory of change model responding to discourses on collaboration, service improvement and outcomes for citizens	The rationale of the local project clashed with the theory of change	The rationale of the project was attuned with the theory of change
EXTERNAL ENVIRONMENTAL FACTORS			
Finance/ESF-LSB funding	£17 million	£151,830 70% financing project manager	£372,220 74% financing project manager and team
Welsh Government staff support	National Project management and evaluation teams	Quarterly communication with ESF-LSB national team No communication with evaluation officer	Quarterly communication with ESF-LSB national team who also undertook a mini-audit of their work
Monitoring	WEFO indicators and theory of change	To address WEFO administrative indicators. No monitoring with regards to project's progress	To address WEFO administrative indicators. Other more fit for purpose indicators were included

Stability of collaborative discourse	The national discourse was changing from innovation through collaboration to outcomes derived from collaboration	The discourse was undecided at the time the project was designed and implemented	The discourse was clearer at emphasising outcomes during design and implementation
COLLABORATIVE OUTPUTS AND OUTCOMES			
Outputs	To be determined by each of the 38 local delivery projects. Very likely to be achieved by all through WEFO indicators	Collaboration agreements, business plans, procurement, shared technical knowledge WEFO targets met	Collaboration agreement, secondments, terms of reference, new funding streams WEFO and other targets were largely met.
Process outcomes (steps on the way to citizen outcomes)	To be determined by each local delivery project. Very likely to be achieved by all	Cost saving across as many local authorities as possible in North Wales	Training youth on employability led to increases in job-search skills and changes in behaviours: improved self-confidence and enthusiasm about work
Citizen outcomes (improvement impacting citizens)	To be determined by each local delivery project. Only a few to achieve these.	Not contemplated, incipient/potential impact	80% of the project's beneficiaries went either employed or in education

Table 2: Outputs in the Caerphilly Passport project

Target	Result (over 2 years)
Create a minimum of 150 work experience opportunities per annum	310 work experience placements
Create a minimum of 25 apprenticeship opportunities per annum	55 apprentice opportunities
Create a minimum of 40 employment opportunities per annum	102 employment opportunities
Reduce the number of people aged 16-24 claiming Job Seekers Allowance by 3% over the lifetime of the project	A reduction of 3.4% claiming JSA due to Passport
Incorporate Targeted Recruitment and Training as a core requirement into at least 10 new public sector contracts (per annum)	Included in 9 new tenders (with others in the pipeline). Also included in 18 contracts as a non-core requirement