



Kent Academic Repository

Hammond, Jonathan, Bailey, Simon, Gore, Oz, Checkland, Kath, Darley, Sarah, McDonald, Ruth and Blakeman, Thomas (2021) *The problem of success and failure in public-private innovation partnerships*. *Journal of Social Policy* . ISSN 0047-2794.

Downloaded from

<https://kar.kent.ac.uk/86530/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1017/S0047279421000192>

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

CC BY-ND (Attribution-NoDerivatives)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

The problem of success and failure in public-private innovation partnerships

Authors

Dr Jonathan Hammond¹ (jonathan.hammond@manchester.ac.uk)

Centre for Primary Care and Health Services Research; Division of Population Health, Health Services Research and Primary Care; School of Health Sciences; Faculty of Biology Medicine and Health; The University of Manchester¹

Dr Simon Bailey

Centre for Health Services Studies; University of Kent

Dr Oz Gore

School of Business; University of Leicester

Prof Kath Checkland¹

Dr Sarah Darley¹

Prof Ruth McDonald¹

Dr Thomas Blakeman¹

Abstract

Public-Private Innovation Partnerships (PPIPs) are increasingly used as a tool for addressing 'wicked' public sector challenges. 'Innovation' is, however, frequently treated as a 'magic' concept: used unreflexively, taken to be axiomatically 'good', and left undefined within policy programmes. Using McConnell's framework of policy success and failure and a case study of a multi-level PPIP in the English health service (NHS Test Beds), this paper critically explores the implications of the mobilisation of innovation in PPIP policy and practice. We highlight how the interplay between levels (macro/micro and policy maker/recipient) can shape both emerging policies and their prospects for success or failure. The paper contributes to an understanding of PPIP success and failure by extending McConnell's framework to explore inter-level effects between policy and innovation project, and demonstrating how the success of PPIP policy cannot be understood without recognising the particular political effects of 'innovation' on formulation and implementation.

Keywords: innovation; public-private partnerships; policy success; policy experimentation

Acknowledgements: We are grateful to the participants for their involvement.

1. Introduction

Public-Private Partnerships (PPPs) are a common yet controversial tool for mobilising a combination of public and private resources towards particular policy objectives. They take many contractual and operational forms but typically involve private sector actors financing the development of assets (frequently infrastructure) linked to a role in the delivery of public services over a given time period. Growing in prominence over recent years, and in the UK particularly since the 1980s (Osei-Kyei and Chan, 2015; Warsen et al. 2018), advocates argue they provide the public sector access to greater resources, increase efficiency, and deliver better value for money (Fischbacher and Beaumont, 2003). Their use is contentious, however; one of the most common forms of PPP are Private-Finance Initiatives (PFI), which enable governments to shift infrastructure expenditure ‘off balance sheet’ and new facilities to be promoted, but have been criticised due to lengthy contracts for private-sector provision of related services (Hodge and Greve, 2007). Local organisations have found the payment of the sometimes very high service charges disadvantageous to achieving financial balance. Moreover, local service provision is exposed to market failures, such as the high profile financial collapse of the PFI provider Carillon, which caused the abrupt halt of construction of two English hospitals to the material detriment of local healthcare demand management.

Our interest in this paper is with a particular branch of PPP: Public-Private Innovation Partnerships (PPIP). PPIPs involve public sector actors collaborating with business with the objective of innovating in the public sector to increase efficiency, respond to particularly challenging problems, and realise new products or services from which private sector collaborators may profit (Brogaard, 2019). Such partnerships are typically of shorter duration (months—several years) than more conventional infrastructure and service provision PPPs, which can last decades (Brogaard, 2019). Public sector innovation policies and initiatives have seen notable growth in popularity in recent decades (Lewis et al. 2019; Osborne and Brown 2011). However, often such initiatives fail to define what innovation is; commonly it is conflated with continuous service change and development, and tends to be treated normatively as an axiomatic good (Osborne and Brown 2011). Nor has the expansion of innovation schemes been matched by developments in the process and practice of their evaluation (Lewis et al. 2019). Indeed, it could be argued that, faced with ‘wicked problems’ (Rittel and Webber, 1973), which are open-ended, inter-connected and without clear pathways to solutions, policy makers reach for apparently simple solutions which cut through the complexity. Concepts such as ‘innovation’, or ‘transformation’ become imbued with almost magical properties in that their invocation obviates the need for justification, specificity or indeed evidence of effectiveness (Pollitt and Hupe 2011).

In this paper, we employ a PPIP case study from the English NHS—NHS Test Beds—to explore the operationalisation of innovation within PPIPs and how this relates to understandings of their success and failure. McConnell’s (2010a) framework for assessing policy success and failure provides a dimensional structure for us to consider success and related PPIP issues more broadly. Motivated by recognition that the last twelve years has seen a push within UK Government towards facilitating local innovation initiatives

(Osborne and Brown 2011), we devote particular attention to the interplay between levels associated with this style of implementation and the organisational factors which can shape prospects for success and failure within ‘policy projects’ (Bailey et al. 2019). This provides scope to think critically about the constructed nature of notions of success and failure in this context as well as the effectiveness and desirability of the current focus on innovation as a means to address hitherto intractable public sector delivery problems. Through our consideration of the interplay between innovation policy agendas at the macro level with more localised levels of implementation, we respond to Ayob et al.’s (2016, p.650) suggestion that research should seek to “unpack the power dimensions” of innovation programmes.

We argue the pursuit of what McConnell refers to as ‘process’ and ‘political’ success at the national policy level can have implications for the feasibility of ‘programme’ success at the local implementation level. Furthermore, ‘magic’ concepts can foreclose political issues at multiple levels. The consequences of this are potentially more significant when it is not government itself, as generally assumed, acting as the most influential policy actor. Lastly, PPIP policies commonly frame programmatic success as partially contingent on the extent to which initiatives ‘spread’ to other areas. Caution should be exercised here as this reveals little about the merit of a policy or initiative and might encourage gaming for reputational benefit.

First, we examine innovation, providing an introduction to public sector innovation policy, and discuss McConnell’s (2010a; 2010b) framework of policy success. Second, we introduce our empirical findings, which examine the evolving experience of a particular UK PPIP initiative. We draw attention to factors we suggest are core components of the ‘policy project’ innovation style—negotiation, iteration, adaptation—and discuss the effects these generate regarding the construction of success and failure at multiple levels.

2. Public sector innovation policy: assessing success and failure

Innovation has become a popular concept with policymakers internationally, characterised by its flexibility, breadth of application, positive associations, and connections with modernist ideas of progress (Osborne and Brown 2011; Ayob et al. 2016; Edler and Fagerberg 2017). Although abundant efforts to classify innovation exist in organisational scholarship (Osborne, 1998; Garcia and Clantone 2002), in policy making there continues to be a lack of precision in the definition and application of innovation, which can lead, in practice, to an ‘anything goes’ implementation approach. Within ‘policy projects’ this might appear a pragmatic means of allowing policy objects to emerge in alignment with local needs and conditions (Bailey et al. 2017). However, this poses problems for defining and measuring success at different times and across different levels of policy making and implementation. As such, innovation warrants classification as a ‘magic’ concept (Pollitt and Hupe 2011). This magical quality can help mobilise support and resources for particular actions (or inactions) and foreclose political debate; appearing simultaneously desirable and unobjectionable and providing latitude for policymakers (Pollitt and Hupe 2011).

Strategies to organise public sector innovation take various forms. Broadly, public service policies and initiatives have shifted from traditional forms of bureaucratic rule-following to

an emphasis on market-based competition and management via performance metrics and targets, commonly associated with New Public Management (Hartley et al. 2013), and a push for greater involvement of private sector actors in public sector innovation. One burgeoning area, which falls under the more general category of ‘collaborative innovation’ (Hartley et al. 2013) or the utilitarian interpretation of ‘social innovation’ (Ayob et al. 2016), is that of PPIP (sometimes referred to as ‘cross-sector innovation’ or ‘inter-organisational innovation partnerships’) where “public and private actors work together to develop innovative solutions targeting the public sector” (Evald et al. 2014, p.34). Research in this area primarily focuses on relationships between partners, how differences between public and private sectors shape joint endeavours, and the importance of factors such as trust, risk sharing and governance processes rather than outcomes and consequences associated with the innovation(s) involved (Brown and Osborne 2013; Evald et al. 2014).

Such evaluation and assessment deficiencies resonate with those recognised as relating to public sector innovation initiatives more broadly. Studies tend to focus on innovation processes rather than assessing the extent to which explicit goals are realised, and when outcomes are assessed these often focus on positives, downplaying negative results or suggesting these demonstrate deficiency with the innovation itself (De Vries et al. 2016). When goals are successfully met a tendency exists for these successes to be treated one dimensionally (Koch and Hauknes 2005), with scant attention to unintended consequences either locally or in the surrounding context. There is also insufficient focus on unsuccessful innovations, their characteristics, or reasons for failure (Koch and Hauknes 2005). Some of these issues relate to problems associated with ‘magic’ concepts; when innovation is understood as inately desirable it is unsurprising other objectives are poorly specified or assessments of their success under-developed.

These issues invite broader questions about how success and failure of innovation in the public sector should be understood, and particularly, what it means for a PPIP to be ‘innovative’ and/or ‘successful’. To understand the potential role of innovation in meeting these goals we must be able to critically interrogate innovation initiatives in ways which transcend local context. To this end, in this paper we apply and develop McConnell’s (2010a; 2010b) framework of policy success within a UK PPIP initiative.

McConnell (2010a) argues that evaluating public policy success is made challenging by partisanship and a tendency for success or failure to be treated as binary, mutually exclusive outcomes, which is reductive and unrealistic. He suggests policies can be conceptualised as comprising three dimensions—process, programme, and politics—and success and failure as a spectrum within each. The ‘process’ dimension refers to policy formulation, whereas ‘programme’ relates to implementation. ‘Politics’ can relate to formulation, implementation, and beyond. Table 1 shows criteria for policy success and failure (at each end of the spectrum) within each policy dimension. Thus, resilient, conflicted and precarious success refer to progressively less successful outcomes. By organising understanding of policy outcomes as combinations of process, programme, and politics a more nuanced, multi-dimensional picture of policy success can be constructed. This may include ‘bundles’ of

potentially contradictory outcomes. For example, a programme may fail to achieve desired outcomes whilst simultaneously proving successful in enhancing leaders' reputations.

[TABLE 1 ABOUT HERE]

3. Methods

This paper draws on an evaluation of NHS Test Bed X (TBX) conducted July 2017–September 2018. The evaluation explored the implementation of TBX with a particular emphasis on the following key criteria: design, adaptation, partnership dynamics, perceptions of benefit, and challenges experienced by stakeholders. Data on the effect of the NHS Test Bed on primary and secondary outcomes were also measured (Lugo-Palacios et al. 2019). The evaluation involved: observations (66 meetings or events, 99 hours); semi-structured interviews with various stakeholders (87 total: 48 with the NHS organisation (20), 'Innovators' (15), Partner and Affiliated organisations (13); 38 with GP practice staff (Practice Managers, Administrators: 19; Clinical: 19), and 1 with an employee of an NHS organisation in a different area (unrelated to TBX) that completed part of the application process for another NHS Test Bed 'wave'; and analysis of documentation relevant to different aspects of TBX (e.g. meeting minutes, agendas, reports, software training guides, messages to primary care stakeholders). Interviews were recorded and transcribed, and contemporaneous field notes written during observations.

Interview transcripts, observational notes, and project documents were imported into NVivo 11 for thematic analysis by JH and SD. This process involved initial coding using the key criteria above as overarching themes. This was supplemented by a 'bottom up' approach that involved 'nesting' codes within overarching themes (e.g. 'Enrolment' within 'Adaptations') as well as developing new high-level themes sitting outside of those, including 'Success claims.' A framework approach was then adopted using McConnell's (2010a) three dimensions so that data extracts could be organised by theme according to their relevance for informing assessments regarding process, programme, and politics success at national and local levels. For example, 'Perceptions of success' sub-codes were relevant to both TBX and the national programme, and these were split between all three dimensions and used heavily; extracts and codes relating to the establishment of TBX were allocated to the process dimension, whereas those pertaining to modifications of the proposed design of TBX were allocated to the programme dimension. This procedure was then further refined with the introduction of McConnell's (2010a) criteria for success within each of three dimensions added (see Table 1). JH led this analysis process, with sense-checking from SB and OG. Extracts were organised according to their applicability to answering questions derived by the criteria themselves (e.g. 'Does this tell us something about whether implementation was in line with objectives?'), both locally for TBX and nationally for the programme at large. Interview extracts below are denoted with a unique

three character interviewee ID code with the month and year of the interview (e.g. [7V3, 08.18]).

4. PPIPs in English healthcare: NHS Test Beds

In this section, we describe our public sector innovation programme—NHS Test Beds—and a particular PPIP project developed as part of the programme: TBX. Before exploring how NHS Test Beds unfolded across both national and local levels, we first describe the context to its implementation.

Since its creation in 2012, NHS England (NHSE), an arm's-length body tasked with overseeing the day-to-day running of the health service and commissioning a range of services including primary care, has taken on de facto responsibility for 'doing' health innovation policy in the English NHS. This was previously the Department of Health's (now Department of Health and Social Care) role, as can be observed in documents such as 'Innovation, health and wealth' (Department of Health 2011). The overarching narrative and objective of innovation has remained largely consistent between government department and arm's-length body: both emphasise the NHS's impressive record developing innovative products and processes, and that continuing to innovate and improving the efficacy and adoption of innovation, is essential to providing good care as well as to benefit the broader economy. Of note, however, is that NHSE has established a new emphasis on the desirability of combining innovations (NHS England et al. 2014), and more broadly, set out a policy direction that diverges from competition between service providers towards a more co-operative system oriented towards place-based planning.

NHS Test Beds

NHS Test Beds is a national policy programme typically involving NHS organisations and private sector organisations (referred to as "innovators") establishing PPIPs. Candidate PPIPs were invited to apply to NHSE for NHS Test Bed status and associated funding, and the first wave of seven Test Beds sites were formally launched in 2016. Taken together the initiatives sought to address various complex health issues and associated challenges (i.e. 'wicked' problems) such as diabetes and self-care or hospital admissions for elderly people with dementia, with patient populations ranging from 1 million–2.5 million. Each Wave 1 NHS Test Bed had an independent local evaluation, and these commonly comprised a process and impact evaluation. Local evaluations were overseen by two national evaluation partners that synthesised results and assessed economic cost.

A concept of explicit importance to the programme is 'combinatorial innovation.' NHSE (2018, p.3) state that this refers to "...different innovations working together rather than, for example, a single blockbuster drug or technology," and, more so, involving "...combinations of types of innovations; for example technology, workforce, new approaches to patient engagement, digital channels for service delivery," rather than multiple different technologies of the same type. The suggestion is that the "...synthesis of different technologies in a joined up way can create synergistic benefits greater than the sum of the parts..." (Galea et al. 2017, pp.4–5) and drive the increases in value sought (NHS England et

al. 2015, p.3). Innovation itself is left undefined save for the particular categories as specified above, and without a foundational definition of innovation it is difficult to define precisely what combinatorial innovation is in this context.

Combinatorial innovation is not a familiar feature of public sector innovation initiatives. Its use here is framed by the idea the NHS has traditionally been poor at implementing new technologies. It promotes an approach connecting the development and introduction of new innovative products with adaptations to processes on the ground. This approach also seeks to realise the specified desirable outcome (both patient and economic benefit) because, it is argued, the era of “silver bullet” innovations is ending (Macdonnell 2015).

In practice, however, it became clear to one interviewee during the early stages of applying to the programme that a combinatorial innovation was also one where an NHS organisation worked in partnership with more than a single ‘innovator’. Ultimately, this interviewee’s organisation abandoned their application because they doubted the feasibility of a project involving additional providers in the time available, noting:

... the chances of success [for the application] seemed very slight at that point when we had that clarification and confirmation of what they were looking for in terms of combinatorial. [4D4, 09.18]

Research on innovation initiatives has identified difficulties embedding them into practice once exposed to organisational realities ‘on the ground’ (De Vries et al. 2016). This challenges the assumption that increasing the number of organisations involved in design and delivery of an innovation in a given setting will increase the chances of success. Even at this very early stage of stakeholder engagement with the programme, it is apparent how the distribution of success/failure across policy dimensions (process, programme, politics) shapes the challenges and possibilities of public policy innovation. Here, a particular rendering of ‘combinatorial innovation’ which equates it with some form of stakeholder cooperation acts to safeguard ‘political’ success (through alignment with broader directions of governance towards co-operation) and deters partners over substantive concerns of ‘programmatically’ success.

NHS Test Bed X

TBX, one of seven Wave 1 NHS Test Beds, was implemented in a single commissioning area in the north of England over a two-year period and involved a partnership between a lead NHS commissioning organisation, a pharmaceutical organisation, and a data analytics organisation. The objective was to improve care for people with one of three long-term conditions. It comprised three components: an IT platform for general practices designed to more effectively manage patients with, or at risk of, a relevant long-term condition, which would also host a bespoke long-term condition risk prediction algorithm; clinical change management and quality improvement using data auditing, feedback, and education sessions; and health monitoring and coaching at a distance using electronic and telecommunications technology. The expectation was that ‘combinatorial benefits’ would be realised by simultaneously implementing these three components within a specific health

care system footprint. Explicit identification of mechanisms of interaction between components was limited, most notably it was intended that potential patients for enrolment into the telehealth element could be identified through the IT platform.

We describe the unfolding of this Test Bed through three themes developed during the analysis process to organise significant events pertaining to PPIP success and failure in temporal sequence: negotiations and delays; design, adaptation, and iteration; outcomes and understandings of success.

Negotiations and delays

Once an initial project proposal was developed between partners and approved by NHSE, a number of problems occurred causing significant delays to the programme. Contract negotiation proved challenging with disagreements over information governance, liabilities, and potential commercial benefits. Implementation eventually began nine months later than anticipated in mid-2017. Incompatible approaches to risk were cited as an issue.

Interviewees from the NHS organisation explained that the organisation was financially unable to adopt certain liabilities relating to the transfer of patient data and there was a sense of surprise from some that the private organisations were reticent to do so. The delay reportedly strained the relationship between some of the senior figures involved in the partnership.

Once contracts and information governance processes were agreed, the next stage involved securing agreement from local general practices (GPs) to have their patients' data used for the programme. A Data Processing Agreement was formulated but the team initially found it hard to get practices to commit. Once the support of the Local Medical Committee (the local representative of the British Medical Association) was secured the majority of practices joined. However, the delay prompted the body with regional oversight for the programme to threaten withdrawal of funding if progress towards implementation was not made, and one interviewee suggested GP practices used participation as leverage to secure better terms with the NHS organisation for an unrelated negotiation. Thus securing 'buy in' from the organisations integral to the project proved problematic, but ultimately all but two practices (5%) were engaged for the duration.

Design, adaptation and iteration

One of the most potentially valuable elements of the project as perceived by many interviewees was the risk prediction algorithm, intended to enable GPs to identify those most at risk of developing one of three long term conditions and offer a proactive service to prevent or delay their development. One interviewee noted that the team recognised implementing this would take time, and to realise 'real world' acceptability they would need to incorporate additional components to demonstrate economic benefits:

I think in an ideal world, yes, we would have just focused on [long term condition development risk prediction algorithm], but in the real world we needed something that would save money quickly. Which is what the [private

sector orgs'] elements were about, so increasing standards, decreasing admissions, therefore saving money. [8T5, 07.18]

Multiple interviewees referenced the scale of ambition and complexity of TBX, which made achieving the intended outcomes challenging.

I don't know whether the other test beds were similarly ambitious, but I guess the thing for me was that it's one thing to be ambitious, isn't it, which is a great thing, but actually I do wonder whether some of that tripped us up a little bit.... And whether or not if they'd have had...looked at maybe just one condition as opposed to a number of conditions which they were looking at, maybe that was an issue... I think if we were going back I'm not so sure we would be quite as ambitious with the programme. [7V3, 08.18]

Relatedly, one interviewee suggested that one of the consequences of this 'combinatorial innovation' design was to make it harder to realise change within the local system:

... if there is one learning from it, it would be around delivering combined innovation and change to the NHS is more challenging than delivering individual components of change, due to the capacity in the system for absorption of change. [2K0, 01.18]

The importance of iteration and development of innovative practices was recognised in the national NHS Test Beds documentation. A number of adaptations were made to the form and staging of the programme, shaped in part by the initial delays incurred. In many cases this undermined the 'test' status of TBX. For example: a planned pilot phase, involving a small group of GP practices providing feedback on the use of the software platform, did not take place. Subsequently, the risk algorithm was significantly scaled back in terms of scope, which meant it was only available for GP practices for a short time (approximately two to four weeks).

Interviewees were concerned by the very limited opportunities for iteration and development of the risk prediction algorithm, in particular:

... you usually build something once, and then you build it again, and then the third time you've really... gotten it right and that's kind of the more general solution. Whereas, I think the Test Beds... my personal opinion is that they... leave time for that second and third phases necessarily. It was, kind of, oh, well, let's build it once and then hopefully it'll just... work and then we can scale it and spread it. But I think that you often need a little bit more iteration to be able to build a thing that actually will scale and spread efficiently. [7X6, 07.18]

Outcomes and understandings of success

The outcome evaluation compared 10-12 months' post-intervention data between TBX's area and another local (without a Test Bed) and found that the Test Bed did not have the expected effect on either primary or secondary outcomes. Consequently, the national

evaluation team decided not to undertake a full cost effectiveness analysis. A more detailed report of TBX's outcome evaluation can be read elsewhere (Lugo-Palacios et al. 2019).

While TBX did not achieve its outcome targets, there were reports of positive impacts (e.g. general practice staff reported improvements identifying and engaging specific patients). Interviewees from the NHS organisation reported unanticipated benefits derived from gathering more up-to-date information about patients with particular long-term conditions in the area, which supported commissioning decisions.

One interviewee talked about success as learning that could inform actions in an unspecified future place and time:

There's no failure in the Test Bed, everything that happened or will happen has got some learning behind it and true success is to learn from them, and then carrying on in an improved manner, or to have that learning available for everyone who wished to implement something similar. So I think that's very important to keep remembering, otherwise by considering some of the challenges as evidence of our failure and not being successful I think can be quite detrimental for everyone involved... [3S8, 01.18]

Relatedly, learning was ascribed to working in partnership with organisations in different sectors, building relationships, understanding alternative perspectives, and creating conditions for future collaboration. For some smaller partner organisations delivering aspects of the programme, it was an opportunity to demonstrate to the larger organisations, particularly the NHS organisation, that they could provide a reliable, quality service to an agreed specification. For the data analytics organisation, having the chance to work with a comprehensive local NHS data set was perceived as a useful experience, lessons from which might benefit future software development. Lastly, learning from TBX was seen as a good basis upon which to apply to the second wave of NHS Test Beds.

5. Process, programme, politics: innovation policy success and failure

In this section, we utilise McConell's (2010b) framework to examine the emergent and constructed definitions of success in NHS Test Beds and PPIP. Subsequently, we suggest how the framework might be further developed.

Process

Process success relates to the extent a policy idea is *formulated* for implementation. Programmes like Test Beds enable relatively loose constellations of ideas and priorities to be tried out in practice. This simplifies policy design; rather than produce detailed 'blueprints' in advance of implementation, local actors are mobilised in the emergent definition *and* implementation of 'bright ideas' (Harrison and Wood 1999). The 'magical' qualities of innovation help to perform this mobilisation. Innovation and Test Beds together provide a symbolic and practical infrastructure to make the policy process 'run'.

At the local level, we might consider formulation as concerning decisions about whether to participate in the programme. As noted, the process of agreeing contracts and liabilities

between partners was costly in terms of time, legal expenses, and goodwill. These delays shaped the form and operation of the programme, and tested partner relationships. Building a sustainable coalition 'on the ground' was tested by resistance from local GPs and securing their involvement cost the NHS organisation financially. Ultimately, sufficient 'buy in' was secured and all elements of the programme launched, albeit with alterations. Locally, TBX was a 'partial' formulation success.

Thinking about process success more broadly, since the global recession of 2008, there has been a marked increase in the number of PPP policies adopted worldwide (Osei-Kyei and Chan 2015). It is often assumed within PPIPs that the public sector lacks innovative capacity and capability because it is rule-bound and bureaucratic, and stands to benefit from the dynamic, competitive forces that animate the private sector. This might be largely mythical (Sørensen and Torfing 2011), however, such axioms are powerful in securing legitimacy for policy proposals. In the absence of foundational definitions of innovation, programmes like Test Beds can accrue a degree of process success simply by 'doing things'.

Our case shows process success can be linked to wider concerns with political accountability. Test Beds was not developed by government but an arm's-length body – NHSE – that has taken responsibility for 'doing' innovation policy, and more broadly has created ambiguity about where the division between the remit of arm's length body and government department lie (Gore, McDermott et al. 2018; Hammond et al. 2018). Test Beds was one aspect of NHSE's 'prescription' for the NHS which they framed as representing the views of health professionals, national leaders, and patient groups. This claim positions NHSE as widening legitimate political participation and simultaneously renders the content and achievement of particular policy goals less politically accountable.

Programme

Programme success is derived from *implementing* policy objectives as intended. It is impossible to draw clear conclusions about the programme success of Test Beds because we lack information about the extent to which Wave 1 Test Beds improved "...patient outcomes and experience of care at the same cost as, or at a lower cost than, current practice, while helping the economy grow" (NHS England 2018, p.3). Emphasis in the available evaluation materials is on the production of learning for future Test Bed waves. The lack of information about outcomes from the individual Wave 1 NHS Test Beds may suggest programme failure, yet the ongoing nature of the programme and emphasis on iterative learning as success acts to defer such an assessment. This implies that programme and process success are closely inter-related with regards the 'test' style of policy making.

This inter-relation is demonstrated by how success is constructed on processual terms. Despite little publicly available information about the evaluation results of the first cohort of Test Beds, in early 2019 a commitment was made to expand the "infrastructure for real world testing" through NHS Test Beds and the creation of "regional Test Bed clusters" to "develop clear operational and business models that are easy for other systems to adopt and adapt, backed by real world data on benefits and costs" (NHS 2019, p.77), and noting that "the primary measure of the success of the Test beds will be the number of other NHS systems that

decide to adopt their models." Programme success, here, is contingent on the extent to which different actors implement its innovation; this implicitly confers legitimacy upon 'the innovation', however, the criteria for success is self-referential.

At the local project level, the desired effects were not realised and consequently its cost effectiveness was not assessed. The programme was delayed, limited in scope in a variety of ways, and the components did not operate concurrently as planned. There were reports of benefit for target groups (primary healthcare professionals and patients) but not sufficient to be reflected in the selected outcomes. As with the national evaluation, success was defined as learning what was and was not working. This broad interpretation of what constitutes success reflects normative qualities associated with innovation and PPIP initiatives more broadly. Taken together, programme success can be considered, at best, precarious.

Politics

The political dimension does not correspond to a particular temporal phase of policy, but concerns political ramifications of a policy for policy makers. McConnell (2010b) emphasises opportunities and benefits, reputation, control, and consistency with broader institutional values as being constitutive of political success.

Our case demonstrates the possibility of government delegating politics and accountability when it comes to large scale health policy development and implementation. NHSE does not have the same partisan or electoral considerations as a government department (Hammond et al. 2018), and one of its concerns is to argue effectively for increased NHS investment from government. This is an inherently political activity, particularly given the background of austerity and because government spending on health comes at the expense of funding other areas.

The Long Term Plan (NHS 2019) claims that NHS Test Beds shows the NHS can 'do innovation' despite the lack of evidence from Wave 1 demonstrating this. The very existence of the programme and its continued implementation is presented as evidence of success, and the label 'innovation' lends legitimacy to such claims. By mobilising innovation in a policy over which it controls the narrative, NHSE is able to shape national agendas; 'doing politics' whilst not being politically accountable to the electorate and with limited parliamentary accountability (Hammond et al. 2018). Nationally, Test Beds is a clear political success.

Locally, TBX was an assemblage of organisations so assessments of political success require attention to its constituents. For the smaller, partially involved organisations it was an opportunity to get a 'foot in the door' with larger organisations and develop a reputation as a reliable provider of NHS services. For these organisations TBX was a political success. The fortunes of partner organisations were more variable, and progress often came with some reputational cost. However, running a Test Bed was perceived as beneficial to chances of succeeding in subsequent applications. This linked to the perceived value of 'doing things'. Political success here is generated simply by continuing to take part in the game.

6. Discussion

We have highlighted the prominence of PPIP, established innovation as a ‘magic’ concept and identified deficiencies in how PPIP policies and initiatives have been assessed and evaluated, which complicates claims to success and value. We engaged these issues via McConnell’s (2010b) framework of policy success and failure in analysing a PPIP case study. Two key contributions can be highlighted: firstly, interplay between levels and dimensions can shape PPIP policies’ prospects for success; secondly, the innovation concept has political implications when mobilised in the formulation and implementation of PPIP policy, and such effects warrant further consideration. We discuss these points below and conclude by considering the politics of innovation and the links between success, scale and spread.

Interplay between levels

Our case study illustrates the importance of interplay between national policy drivers and local policy realisation in PPIP. This policy was presented as a means for developing and spreading ‘combinatorial’ innovation, intended to generate novel products whilst changing practices on the assumption this would create synergistic benefits. This novel construction, which enhances the ‘magic’ effect of innovation upon policy, facilitates both process and political success at the national/policy level. In practice, budding NHS Test Bed teams needed to demonstrate these principles in their applications, and include multiple private sector partners, in order to be selected. Given the complexities involved in delivering a ‘combinatorial’ programme, this suggests that the very qualities contributing to process and political success at the national level (i.e. the association with the ‘magic’ of innovation) can undermine local projects attempting to realise programme success and, by extension, chances of national programme success. In other words, the very features of the policy which gave it process and political success at national level mitigate against local and national programme success. This resonates with issues relating to PPPs more broadly, highlighted at the start of the paper, in that they can be attractive to policy makers in the short term yet come with significant implications regarding risk sharing, governance, and costs over longer time horizons (Hodge and Greve, 2007).

Our findings suggest that PPIP policies involving local programmes, regardless of sector, should provide space for partnership development, and for the emergent collective identification of what is effective in the specific local operating context, rather than imposing national rules (potentially driven by political requirements) which may establish dynamics that can militate against success. Subsequent attempts to spread these initiatives may benefit from additional depth of understanding about what local adaptations proved necessary in order to inform the process of tailoring to different specific contexts.

Developing frameworks for policy success

Our analysis demonstrates value in using McConnell’s (2010b) framework to study policy from the top down and consider dimensions of success between levels. To our knowledge this is the first time that it has been used in this way. This does not, however, address existing issues with the approach including the uncertainty around the question of ‘success

for whom?’ (Marsh and McConnell 2010). Our case shows that different organisational actors can be assessed as achieving different degrees of success, across the programme and politics dimensions in particular. This is partly a consequence of extending the framework to encompass a focus on projects as well as policy, and there is no straightforward solution to this beyond recognising and exploring differences between stakeholders in their experiences, dynamics, outcomes and opportunities. This is itself a useful exercise, however, because it allows PPIP projects to be ‘unpacked’ to facilitate greater understanding about how certain policies or approaches work or fail to work as intended.

Of relevance to PPIP policy is the process success criterion of ‘symbolising innovation and influence’. While this criterion arguably represents more a statement of fact (i.e. policies that symbolise innovation are likely to get off the ground) than a normative aspiration, it is important to consider this criterion as potentially problematic. As demonstrated, ‘innovation’ is a powerful and underspecified organising concept that can have a range of implications across programme and politics dimensions. Future studies should be sensitive to this.

PPIP and ‘policy tests’ are becoming increasingly dominant as approaches to public service orchestration become more ‘projectified’ (Hodgson et al. 2019). It is important to consider how the logics and techniques of projects might have a determining influence upon the kinds of problem considered suitable for programmes such as Test Beds, as well as the emergent shaping and measurement of particular initiatives. This is in part an issue of temporality – an essential feature of projects as temporary forms of organisation. We have noted the effects of this within our case. Connected to the issue of time is the instrumental rationality which projects impose on their protagonists; that is, the need for measurable outcomes mobilises a search not for what might or could work, or work most effectively, but what can be made to work within the time available (Bailey et al. 2017; Bailey et al. 2019; Goff et al. 2021). This links to the self-referential nature of success noted above and may be somewhat more modest than the claims made during the tendering stage to meet the expectations for a successful application. As policy projects generally rely upon a champion or champions equipped with local knowledge and resources to get them off the ground and make them work, the performative nature of the project logic combines with local power and reputational interests to shape the meaning and measurement of ‘innovation’ in such cases.

The politics of innovation and public-private partnerships

The idea of partnerships in social policy, like innovation, enjoys a certain degree of axiomatic desirability. As Rummery (2002, p.243) notes “who could possibly object to partnership as a concept?”, yet partnerships between private and public organisations are inherently political and can reinforce existing power inequalities. A notable trend in many developing countries is an increase in PPPs in education provision. This is often framed as a straightforward technical solution to a resource problem, yet the network of policy entrepreneurs representing private interests driving this development do so on the basis of a shared understanding of the desirability of promoting private sector development through

education (Verger, 2012). Such issues are not neutral or one dimensional. Public-private dynamics warrant particular consideration in PPIP when an explicit objective is the generation of new sources of economic value through the development of novel products and practices, and the expansion of those deemed successful. It is important to explore the political implications of such arrangements in relation to assessments of policy success because both the innovation concept and associated arm's length governance, where this is found, have the potential to foreclose or stifle the political dimension (Hammond et al. 2018). This highlights a deficiency with McConnell's (2010b) framework which assumes government occupies the central policy actor role. We have highlighted the relevance of exploring the status of policy makers and the dynamics of the system within which they operate because of the potential implications this has for policy and programme success.

A key question that our case prompts is: why is a policy that has not demonstrated clear benefits being rolled out more broadly? Our analysis suggests the innovation concept provides a buffer, so programmatic success becomes less essential than it would otherwise be. All outcomes are badged as 'learning' to inform some future unspecified success. The appeal to policy makers here is clear as innovation policies represent a source of 'easy' reputational success due to their legitimacy and reduced delivery pressures. In our case it is NHSE, an arm's length body, that has seemingly adopted responsibility for orchestrating innovation policy and operates more broadly as a health service meta-governor (Hammond et al. 2018). NHSE was specifically created with the objective of removing 'political interference' from NHS oversight, but has taken a pro-active role re-engineering the balance between hierarchies, markets, and networks across the system. The dynamic between arm's length body and government continues to evolve. NHSE has at times appeared to provide a useful insulator from reputational damage for government ministers when problems with the health service have occurred (Hammond et al. 2018); more recently reports from government sources suggest that NHSE possesses too much power and this should be curtailed (West 2020). Both provide grounds to question the efficacy of the framework intended to oversee and hold NHSE activities to account.

Success, scale & spread

It is common for PPIP policies to present the extent to which an innovation initiative spreads to other areas as a yardstick for programmatic success. The very language of 'test bed' positions the programme at the 'initiation' phase of a longer innovation trajectory (Ven et al. 2008). We argue that this might be problematic for public sector innovation policies and caution should be exercised. In countries with established public sector systems, policy 'sedimentation' can occur whereby layers of policy consequences build up and weave into rich inter-organisational histories and varied interests (Hammond et al. 2017; Gore, Hammond et al. 2018). Such systems are not composed of uniform organisations (or broader networks or collectives) that are equally receptive to the successful introduction of particular innovations. These sub-systems are relationally constituted, unique, pressurised, and constantly evolving (Hammond et al. 2017). Given this, programmatic success might come to reflect those initiatives that are most transposable and these may or may not bear particular connection to those which are of most potential value to system stakeholders. At the

national/policy level, in this case, programmatic and political success rests on evidence that innovations are spreading, rather than on the outcomes associated with them. At the local level, PPIP or other organisational entities stand to achieve political success if their innovation spreads to other areas but this creates incentives for local actors to make bold claims about the coherence, transposability, and performance of their initiatives. Future research might usefully focus specifically on some of these issues.

Limitations and broader significance

It is important to recognise the specificities of the NHS Test Beds case and consider what implications there may be for extending the conclusions of analysis to other PPIPs. NHS Test Beds involves a national policy driven by an arm's length body, in collaboration with a variety of other organisations, and the development and implementation of numerous innovations at the local level. Certainly not all PPIPs exhibit these qualities and thus our particular findings should not be expected to transfer to other PPIPs directly. Our analysis frame using McConnell's (2010a) process, programme, and politics dimensions is, however, applicable to all PPIPs, and indeed PPPs more generally. All PPIPs have the potential for inter-dimensional dynamics to affect the prospects for success in a given dimension, and these may or may not be further complicated by the inter-play of levels (i.e. national, regional, local) depending on the case. Furthermore, our case study highlights the inherent challenges of mobilising innovation through PPIPs as a means of addressing wicked problems. This is because such challenges are by definition open-ended, and inter-connected, and PPIPs tend to involve acutely time-limited projects with multiple actors potentially pursuing success in different dimensions. Our findings are therefore relevant to other similarly 'projectified' policy initiatives (Hodgson et al. 2019), whether or not they are badged as PPIP.

Finally, returning to the question of 'success for whom?', our findings show that our case study, and PPIPs more generally, should be understood as not delivering universal 'goods' for stakeholders involved. Benefits are distributed unpredictably, unevenly, and do not necessarily align with resources stakeholders invest. In the case of NHS Test Beds and TBX, policy makers derived political success irrespective of local outcomes. However, local programme success was limited for the main NHS and private sector organisations involved, partially as a consequence of the scale of ambition involved in the time available, but smaller provider organisations perceived some reputational (political) success as a consequence of their involvement. This resonates with research on policy piloting, which have been described as a form of 'government at a distance' (Foucault, 1991), permitting multiple definitions of success among programme designers and implementers (Bailey et al. 2017; Bailey et al. 2019). This highlights perhaps the most valuable aspect of McConnell's (2010a) framework when applied to PPIPs: it enables the articulation of the uneven topography of success, and the conditions in which certain stakeholders may feel they receive a 'return' on investment even in the absence of overall programmatic success. More broadly, an uneven distribution of successful outcomes signals the potential for programme fragmentation, reproduction of existing divisions and inequalities and increasing heterogeneity within and between places among service providers. Such issues may be

concealed by claims of success by some stakeholders in one dimension, and the mobilisation of innovation in policies and partnerships provides fertile conditions for this to occur. The implication of this are inherently political and warrant attention.

References

- Ayob, N., Teasdale, S. and Fagan, K. (2016). How social innovation 'came to be': tracing the evolution of a contested concept. *Journal of Social Policy*, 45(4), pp.635–653.
- Bailey, S., Checkland, K., Hodgson, D., McBride, A., Elvey, R., Parkin, S., Rothwell, K. and Pierides, D. (2017). The policy work of piloting: mobilising and managing conflict and ambiguity in the English NHS. *Social Science & Medicine*, 179, pp.210–217.
- Bailey, S., Hodgson, D. and Checkland, K. (2019). Pilots as projects: policy making in a state of exception. In D. Hodgson et al., eds. *The projectification of the public sector*. Abingdon: Routledge, pp. 130–148.
- Brogaard, L. (2019). Innovative outcomes in public-private innovation partnerships: a systematic review of empirical evidence and current challenges. *Public Management Review*, 0, pp.1–23.
- Brown, L. and Osborne, S. (2013). Risk and innovation. *Public Management Review*, 15(2), pp.186–208.
- De Vries, H., Bekkers, V. and Tummers, L. (2016). Innovation in the public sector: a systematic review and future research agenda. *Public Administration*, 94(1), pp.146–166.
- Edler, J. and Fagerberg, J. (2017). Innovation policy: what, why, and how. *Oxford Review of Economic Policy*, 33(1), pp.2–23.
- Evald, M.R. et al. (2014). Reviewing cross-field public private innovation literature: current research themes and future research themes yet to be explored. *International Public Management Review*, 15(2), pp.32-57.
- Fischbacher, M., Beaumont, P.B. (2003). PFI, public-private partnerships and the neglected importance of process: stakeholders and the employment dimension. *Public Money & Management*, 23(3), 171–176.
- Foucault, M. (1991). Governmentality. In G. Burchell, C. Gordon, & P. Miller, eds. *The Foucault effect: studies in governmentality*. Chicago: University of Chicago Press, pp. 87–104.
- Galea, A., Hough, E. and Khan, I. (2017). *Test Beds the story so far*. London: NHS England.
- Goff, M., Bailey, S., Hodgson, D., Bresnen, M., Elvey, R., Checkland, K. (2021) Ambiguous workarounds in NHS policy piloting: tensions, trade-offs and legacies of organisational change projects. *New Technology, Work & Employment*, In Press

- Gore, O., McDermott, I., Checkland, K., Allen, P., Moran, V. (2018). Discretion drift in primary care commissioning in England: Towards a conceptualization of hybrid accountability obligations. *Public Administration*, 98(2), pp.291-307
- Gore, O., Hammond, J., Bailey, S., Checkland, K. and Hodgson, D. (2018). Not every public sector is a field: evidence from the recent overhaul of the English NHS. *Public Management Review*, 21(4), pp.559-580.
- Hammond, J., Lorne, C., Coleman, A., Allen, P., Mays, N., Dam, R., Mason, T. and Checkland, K. (2017). The spatial politics of place and health policy: exploring Sustainability and Transformation Plans in the English NHS. *Social Science & Medicine*, 190(Supplement C), pp.217–226.
- Hammond, J., Speed, E., Allen, P., McDermott, I., Coleman, A. and Checkland, K. (2018). Autonomy, accountability, and ambiguity in arm’s-length meta-governance: the case of NHS England. *Public Management Review*, 21(8), pp.1148–1169.
- Harrison, S. and Wood, B. (1999). Designing health service organization in the UK, 1968 to 1998: from blueprint to bright idea and ‘manipulated emergence’. *Public Administration*, 77(4), pp.751–768.
- Hartley, J., Sørensen, E. and Torfing, J. (2013). Collaborative innovation: a viable alternative to market competition and organizational entrepreneurship. *Public Administration Review*, 73(6), pp.821–830.
- Hodge, G.A. and Greve, C. (2007). Public–private partnerships: an international performance review. *Public Administration Review*, 67(3), pp.545–558.
- Hodgson, D. et al. (2019). *The projectification of the public sector*. Abingdon: Routledge.
- Koch, P. and Hauknes, J. (2005). *On innovation in the public sector*. Oslo: NIFU STEP.
- Lewis, J., McGann, M. and Blomkamp, E. (2019). When design meets power: design thinking, public sector innovation and the politics of policymaking. *Policy & Politics*, pp.1–20.
- Lugo-Palacios, D.G. et al. (2019). The impact of a combinatorial digital and organisational intervention on the management of long-term conditions in UK primary care: a non-randomised evaluation. *BMC Health Services Research*, 19(1), p.159.
- Macdonnell, M. (2015). Innovation: ‘silver bullet’ solutions are out and Test Beds are in. *Health Service Journal*. [online]. Available from: <https://www.hsj.co.uk/comment/innovation-silver-bullet-solutions-are-out-and-test-beds-are-in/5085535.article> [Accessed March 2, 2020].
- Marsh, D. and McConnell, A. (2010). Towards a framework for establishing policy success. *Public Administration*, 88(2), pp.564–583.
- McConnell, A. (2010a). Policy success, policy failure and grey areas in-between. *Journal of Public Policy*, 30(3), pp.345–362.

- McConnell, A. (2010b). *Understanding policy success: rethinking public policy*. Houndmills: Macmillan International Higher Education.
- NHS. (2019). *The NHS long term plan*. London: NHS England.
- NHS England et al. (2014). *Five year forward view*. London: NHS England.
- NHS England et al. (2015). *Real world testing of 'combinatorial innovation.'* A global invitation to innovators. NHS England.
- NHS England. (2018). Test Beds programme: information to support the launch of wave 2. [online]. Available from: <https://www.england.nhs.uk/wp-content/uploads/2018/04/test-beds-supporting-information-wave-2-competition.pdf>.
- Osborne, S.P. and Brown, L. (2011). Innovation, public policy and public services delivery in the UK. The word that would be king? *Public Administration*, 89(4), pp.1335–1350.
- Osei-Kyei, R. and Chan, A.P.C. (2015). Review of studies on the critical success factors for public–private partnership (PPP) projects from 1990 to 2013. *International Journal of Project Management*, 33(6), pp.1335–1346.
- Pollitt, C. and Hupe, P. (2011). Talking about government: the role of magic concepts. *Public Management Review*, 13(5), pp.641–658.
- Rittel, H.W.J., Webber, M.M., 1973. Dilemmas in a general theory of planning. *Policy Sciences*, 4(2), pp.155–169
- Rummery, K. (2002). Towards a theory of welfare partnerships. In G. Caroline, P. Martin, & K. Rummery, eds. *Partnerships, New Labour and the governance of welfare*. Bristol: Policy Press, pp. 229–246.
- Sørensen, E. and Torfing, J. (2011). Enhancing collaborative innovation in the public sector. *Administration & Society*, 43(8), pp.842–868.
- Ven, A.V. de et al. (2008). *The innovation journey*. Oxford: Oxford University Press.
- Verger, A. (2012). Framing and selling global education policy: the promotion of public–private partnerships for education in low-income contexts. *Journal of Education Policy*, 27(1), pp.109–130.
- Warsen, R., Nederhand, J., Klijn, E.H., Grotenbreg, S., Koppenjan, J. (2018). What makes public-private partnerships work? Survey research into the outcomes and the quality of cooperation in PPPs. *Public Management Review*, 20(8), 1165–1185.
- West, D. (2020). Take NHS England's power - but don't give it to ministers. *Health Service Journal*. [online]. Available from: <https://www.hsj.co.uk/comment/take-nhs-englands-power-but-dont-give-it-to-ministers/7026887.article> [Accessed March 2, 2020].

	Success			Failure				
<i>Process</i>	Preserving government policy goals and instruments	Resilient success	Conflicted success	Termination of government policy goals and instruments				
	Conferring legitimacy on the policy			Irrecoverable damage to policy legitimacy				
	Building a sustainable coalition			Inability to produce a sustainable coalition				
	Symbolizing innovation and influence			Symbolizing outmoded, insular or bizarre ideas, seemingly oblivious to how other jurisdictions are dealing with similar issues				
	Opposition to process is virtually non-existent and/or support is virtually universal			Opposition to process is virtually universal and/or support is virtually non-existent				
<i>Programme</i>	Implementation in line with objectives			Precarious success	Precarious success	Implementation fails to be executed in line with objectives		
	Achievement of desired outcomes					Failure to achieve desired outcomes		
	Creating benefit for a target group					Damaging a particular target group		
	Meets policy domain criteria					Clear inability to meet the criteria		
	Opposition to program aims, values, and means of achieving them is virtually non-existent, and/or support is virtually universal					Opposition to program aims, values, and means of achieving them is virtually universal, and/or support is virtually non-existent		
<i>Politics</i>	Enhancing electoral prospects or reputation of governments and leaders					Precarious success	Precarious success	Damaging to the electoral prospects or reputation of governments and leaders, with no redeeming political benefit
	Controlling policy agenda and easing the business of governing							Policy failings are so high and persistent on the agenda, that it is damaging government's capacity to govern
	Sustaining the broad values and direction of government							Irrevocably damaging to the broad values and direction of government
	Opposition to political benefits for government is virtually non-existent and/or support is virtually universal							Opposition to political benefits for government is virtually universal and/or support is virtually non-existent

Table 1: Spectrum of success (and criteria for success and failure) within policy, programme, and politics dimensions of policy (Adapted from McConnell (2010b))