


Evaluation in reinforcing and resisting hierarchical relations between state and civil society

Sioned Pearce  | Helen Blakely | David Frayne | Ian Rees Jones

Wales Institute of Social and Economic Research and Data (WISERD), Cardiff University, Cardiff, UK

Correspondence

Sioned Pearce, Wales Institute of Social and Economic Research and Data (WISERD), Cardiff University, 38 Park Place, Cardiff CF10 3BB, UK.
Email: pearces11@cardiff.ac.uk

Funding information

Arts and Humanities Research Council;
Economic and Social Research Council, Grant/Award Number: ES/L009099/1

Abstract

Here we present a critical exploration of evaluation as a concept within a state-led social policy programme. Studies critiquing this type of evaluation often assume its purpose is to provide knowledge and understanding of a given social policy, and its relative impact upon the social issue towards which it has been directed. However, drawing on the accounts of 25 community development workers gathered over the course of a 17 year state-led, anti-poverty programme (2001–2018), and building on existing critique of evaluation methodologies, we argue that evaluation is also instrumental in the reinforcement of hierarchical power relations between state and civil society. To develop this argument, evaluation is discussed in three related ways pertaining to hierarchy: (a) firstly, as a means of defining and ultimately producing (contested) constructions of value; (b) secondly, as a mechanism for securing forms of vertical accountability; (c) and finally, through its construction as a *lost saviour*: an entity with untapped potential for safeguarding the integrity of an initial political ideology. In this way, narratives from those working on the ground extend our understanding of the complexities and dualities embedded within evaluation. In light of this analysis, we argue for a more inclusive approach to evaluation practices, and the development of alternative heterarchies in the

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Social Policy & Administration published by John Wiley & Sons Ltd.

evaluation of social policy premised on processes of co-production and collaboration.

KEYWORDS

civil society, community development, evaluation, hierarchy, the state

1 | INTRODUCTION

The past 10 years have seen a growth in critique of evaluation as a concept in the field of social and public policy (Lamont, 2012, 2017; Lamont, Beljean, & Clair, 2014; Prior, 2008; Timmermans & Epstein, 2010). Definitions of value constructed by institutions of governance are understood to compound a growing sense of inequality (Lamont, 2012), and criticism is levelled at evaluation (and programme) design for failing to accurately capture the full impact of policy and the related (mis)use of evaluation findings. Further to this, a high degree of resistance to evidence amongst policy makers is thought to negate the unprecedented amount of information available to them, making it 'the best of times and the worst of times' for policy design (Peters, 2018, p. 3). An example of this contradiction is the recent closure of the Department for International Development in the United Kingdom, which had demonstrated a significant degree of openness, rigour, and innovation with respect to evaluation. Yet its closure was justified by the argument that government must be 'fearless [in] utilising data analytics specialists, to more rigorously evaluate policy successes and delivery failures' (Gove, 2020).

Defining value through evaluation, and particular evaluation methodologies, can be considered a form of governmentality whereby state power is ubiquitous within, in this case, a given social policy (Foucault, 1980) and therefore, embedded within the process of evaluation itself. In this way, Larner and Butler (2005) argue, evaluations are not simply neutral tools but governmental techniques that 'represent and help constitute governmental spaces and subjects in particular forms' (2005, p. 87). Such analyses are positioned in direct opposition to the new public management (NPM) and to a lesser extent public administration, as cultural phenomena within public services with strong implications for evaluation practices (Espeland & Sauder, 2007; Newman, 2005; Osborne, 2006; Power, 1997). Of concern, here are those elements of NPM that emerge in the evaluation of a state-led social policy programme, namely: the adoption of private-sector management approaches within public sector operations; the distancing between policy makers and policy implementation; the decentralisation of management authority within public agencies; and the subsequent individualisation of responsibility for policy delivery (Osborne, 2006, p. 379; Pollitt, 1995, p. 134).

We extend these arguments by drawing on empirical data, which suggests that the process of evaluation does not simply identify success, attribute accountability for failure, or prescribe solutions; but can also serve to reinforce hierarchical relations between state and civil society, or the hierarchy of command in governance (Jessop 2002). Of particular interest is the conversion of value into metrics, a key feature of NPM, through processes of output or indicator-based performance management and accountability. Widely argued to devalue public services, this aspect of NPM drives practices on the ground and has become a key locus for contemporary critical assessments of evaluation (Bevir, Needham, & Waring, 2019; Lowe & Wilson, 2015). Building upon existing analyses of evaluation and its role within social policy programmes across Western Europe and North America (Allen, Needham, Hall, & Tanner, 2018; Bovaird, 2012; Parr & Churchill, 2019; Richardson, 2013; Vedung, 2017), we argue that while a preoccupation with performance hinged on NPM can impede the development of policy and practice (Peters, 2018, p. 140), we should not assume that this is its sole purpose, particularly from the perspective of those on the ground. Rather, analyses of evaluation can make visible those frameworks of formal bureaucratic procedures, in which hierarchical relations are embedded and enacted (Roethlisberger & Dickson, 1941); here, information-gathering systems

with a focus on outputs are understood to be central to the production and circulation of power. In this vein, we highlight the way prevailing NPM evaluation techniques drive the individualisation of responsibility for programme success or failure, with implications for understandings of (self) worth amongst practitioners. This approach differs from a number of key studies critiquing evaluation within social policy where the focus is relative impact upon the social issue towards which it has been directed (Larner & Butler, 2005; Espeland & Sauder, 2007; Freeman & Maybin, 2011; Richardson, 2013; Allen et al., 2018; Parr & Churchill, 2019).

At this point it is useful to introduce alternative approaches to evaluation that have gained traction, albeit for the most part within academia and civil society, and which are indicative of the evolution of interventions beyond those produced through the NPM. These models allow us to conceive of evaluation practices and processes that purposively incorporate the knowledge and experience of those working on the ground. Specifically, authors focusing on Realist(ic) and Developmental evaluation emphasise the importance of policy theory, policy embeddedness, policy activity and open systems of evaluation in the case of the former (Pawson & Tilley, 2004) and policies that support social innovation and adaptive management in the case of the latter (Patton, 2011). In its aim to expand evaluation beyond a focus on auditing outputs, a Realist(ic) perspective rests on the assumption that the purpose of evaluating is to improve policy 'through theory testing and refinement' (Pawson & Tilley, 2004, p. 9). Specifically, Pawson and Tilley ask: 'what works for whom, in what circumstances and in what respects, and how?' (ibid, 2004, p. 2). We will return to this approach and its implications in the concluding discussion as a means of developing our findings. Most significantly, we highlight the need for meaningful recognition of practitioner knowledge and experience in both policy development and its interrogation through evaluation practices, as noted in the democratic experimentalism literature:

The commitment to testing belief against experience, freedom to criticize established views, transparency and free access to information, and a sense of collaboration among peers. (Sabel & Simon, 2017, p. 3)

Thus, we hope to clarify the purpose of the evaluation process beyond an understanding of what works (for whom and in what circumstances) towards addressing the question of 'knowledge for whom and for what?' (Burawoy, 2004, 2012), and the attendant implications for the production of value through policy.

The remainder of the paper is divided into four parts. The first details the background and method for this empirical study, outlining the context for the research study and its design. The second part of the paper provides an overview of the external and internal evaluations of Communities First. The third, drawing on empirical data collected with community development workers, focuses on findings. In this part, evaluation is discussed in three related ways pertaining to hierarchy: (a) firstly, as a means of defining and ultimately producing contested constructions of value; (b) secondly, as a mechanism for securing forms of vertical accountability; (c) and finally, as a *lost saviour*—an entity with untapped potential for safeguarding the integrity of an initial political ideology.¹ These findings attend to the gap in knowledge on perceptions of evaluation, its purpose and the motivations driving it, culminating in a discussion of the tensions emerging from perceptions of power through categorisation and legitimisation in evaluation. Subsequently, we reflect on the privileging one type of knowledge over another (Lamont et al., 2014) and why it matters for wider issues of hierarchy, power and sustainability in community development. Finally, the concluding discussion revisits our understandings of evaluation in light of our analysis with a view to offering alternative approaches.

2 | THE STUDY

The social policy examined here is the Welsh Government's flagship community development programme designed to address poverty in Wales, Communities First. Communities First ran for 17 years across Wales under a devolved Welsh Government, aiming first to *tackle* and later to *address* poverty and deprivation as defined by the eight

domains of the Welsh Index of Multiple Deprivation. Communities First has been chosen as the site for a critical exploration of evaluation because of its longevity, its community development focus and the extent of its scrutiny through state-led evaluation over 17 years. This community development programme with historic roots in civic activism and participatory democracy forged, in part, through resistance to conventional forms of state intervention, also provides fruitful ground for better understanding reinforcement of and resistance to hierarchical relations through evaluation. Designed to empower people living in the most deprived areas of Wales, this programme was originally conceived as a grass roots approach to tackling the issues generated by poverty, as opposed to an attempt to address the structural drivers of poverty itself (Adamson, Dearden, & Castle, 2001). Delivered by community development workers in the most deprived localities² of Wales, it was initially implemented in approximately 150 small geographical areas, averaging 4,000 in population. The programme was rolled-out on the ground through partnership bodies made up of representatives of the statutory, community, business and/or civil society in each small area, each of which was coordinated by one or more community development professional(s). Costing around £300 million (National Assembly for Wales, 2017), Communities First was subject to considerable scrutiny, not least through the Welsh Government's own internal and external evaluations. The internal evaluation is the focus of this empirical study. Over time, this form of evaluation evolved from a broadly qualitative approach to an increasingly metric-driven quantitative model, reflecting a programme-wide tightening of state directive. The practices of the internal evaluation came to emphasise outputs over outcomes in keeping with the NPM approach. The termination of Communities First was announced in 2017 on the basis that '... poverty remains a stubborn and persistent challenge' (Carl Sargeant, 14th February 2017). Project funding ended in April 2018.

Taken as a whole, the dataset for our empirical study consists of a total of 55 qualitative semi-structured interviews carried out between 2009 and 2018 (with community development workers, policy-makers, civil servant, statutory sector workers, civil society representatives, volunteers, local residents, academics), ethnographic fieldnotes made during an internship with the Welsh Government in 2009, and an analysis of 16 key documents. The research was an independent, academic study of the programme funded by UK research councils. For this paper, we focus on the qualitative interview data collected with 25 community development workers, however, interviews with other representatives have been used as valuable points of comparison and contextualisation. When contacted, it was explained to the participants that the purpose of the research was to explore their experiences of the Communities First programme. During the interviews, which were open-ended and built around employment histories, community development workers often took the time to detail at length the significance of evaluation practices and processes for their everyday working lives, without being prompted. This interview data was gathered from three Communities First case study areas representing valley towns, rural, and urban geographies in two phases during 2009–2010 and 2017–2018, both of which were key periods of change for the programme and its evaluation. The latter phase of data collection, at the programme's end, signified a crisis point for community development workers as the structures within which many had worked for nearly 20 years were being dismantled. However, this period was also characterised by a valuable moment of reflexivity, giving workers the space to talk about their role in the programme and to contest the idea that it, and by extension they, had *failed* to counteract the forces driving poverty in Wales. In this newly created reflexive space it is possible to identify accounts of the way in which evaluation has reproduced hierarchies.

2.1 | Background to programme evaluations

At this point, it is important to distinguish between the nine externally commissioned, programme evaluations carried out at various stages during Communities First's lifespan and the internal, ongoing government led evaluation carried out annually over the programme's 17 years. The external can be seen as typical of social policy evaluation research and analysis (Vedung, 2017) while the internal has been less well explored and is the focus of this paper. The crucial difference lies in the object under scrutiny and by extension the purpose of the

scrutiny. In this case, and in broad terms, the external evaluations scrutinised government policy while the internal evaluation scrutinised the work of those delivering it. Significantly, what is termed internal evaluation here was commonly referred to as 'monitoring' by Welsh Government officials and 'evaluation' by community development practitioners. We have chosen to adopt the term 'internal evaluation' recognising that there are diverse forms of categorisation going on here. To expand on this, some background to the external evaluation of Communities First is now given.

2.1.1 | External evaluation

External evaluations of Communities First, largely commissioned by the Welsh Government, from 2001 onwards are useful markers of continuity and change in the programme's ethos. Taken as a whole they reveal a shift in the underpinning principles surrounding the programme, most notably from *community led* activities designed to empower local people and build capacity within communities, to a *community focused* approach designed to generate particular outputs and thereby better outcomes for communities (Pearce, Sophocleous, Blakely, & Elliott, 2020). Crucially the latter was associated with a desire on the part of government to be in a position to better measure the impact of the policy (Adamson & Bromilley, 2008; Cambridge Policy Consultants, 2006; Coleman, 2009; Hincks & Robson, 2010; Ipsos MORI and Wavehill Consulting, 2015; National Assembly for Wales, 2003; National Assembly for Wales, 2017; Scorrer & Adamson, 2007; Wavehill Consulting, 2007; Welsh Government 2006). External evaluations had consistently identified the impossibility of quantifying the (typically economic) impact of community development work as deeply problematic. For example, in September 2006 Cambridge Policy Consultants published an interim evaluation of Communities First. This document was positive about the progress of partnerships and the role of Welsh Ministers, among other things, but argued for a more systematic approach to measuring change through a community led approach:

The [Welsh Assembly Government] ... should introduce a common framework for local evaluation... The systems should be designed to record evidence of the process and outcomes of change. Partnerships should work with a compulsory set of core indicators that would inform the WAG about their progress, however the system should contain many more indicators that would allow each partnership to record its own success measures and stories of achievements. (Cambridge Policy Consultants, 2006, p. 112)

Three years later, AMION Consulting's Evaluation (2011) was broadly positive in contrast to the more critical, independent academic assessment of Adamson and Bromilley (2008), but again concluded that the challenge of measuring economic value remained:

... it is not possible to undertake a detailed quantified "top down" analysis of the value for money of the CF programme. (AMION et al., 2011, p. 110)

In 2015, the Ipsos MORI evaluation raised the same difficulty of quantifying the impact of the programme:

... no firm conclusion is drawn about the extent to which Communities First is achieving its aims. Indeed, an assessment of its outcomes is likely to be hampered by the availability of beneficiary data and robust monitoring information and by the design of an area based intervention to achieve individual-level change. (Ipsos MORI, 2015, p. 88)

Over time, the programme changed but unsurprisingly, the issues associated with evaluating the impact of hyper-localised, 'soft' approaches to addressing poverty within multiple, small areas persisted. During this time,

we can also see an increasing tendency by government to seek 'returns' on policy 'investments', compounding a perception on the ground of a sharp incongruity between what was being achieved and what was being measured.

2.1.2 | Internal evaluation

In contrast, the long term, internal evaluation of the programme was led by a research team of civil servants within the Welsh Government. Between 2001 and 2007, annual internal evaluation reports were written and submitted by Communities First Partnerships to local authorities who in turn submitted the reports to Welsh Government. Reporting templates were simple Word documents with open text sections for entering data on 'progress' and 'support'. In this reporting template, partnerships decided broadly what should be recorded, and relatedly determined how success should be measured and reported. This process was underpinned by an action research approach to community development practice. The following quote from a Welsh Government civil servant sheds some light on the way this early evaluation process worked:

I remember that action research period where the Head of Communities First at that time ... had sort of six weekly or two monthly meetings with the research team (in the Welsh Government) who would go off and listen to, talk to the funded communities and come back and say this is what the funded areas are doing, these are the sort of case studies, these are the problems that they've got, these are the constraints, these are the solutions and then use that to influence the way the programme was working. So that's a much more operational type use of evidence ... because it was a constant process, but also it sort of addressed and brought in some of these wider ideas through the analysis that those researchers that we brought in could bring, you know their understanding of how community participation works for example. (Welsh Government Civil Servant 8.3, January 2017)

However, from 2008 the process was changed by the Welsh Government. The reporting structure became metric, through the addition of a layer of reporting to be completed by local authorities using numeric data submitted by partnerships each year. Word templates were replaced by Excel spreadsheets and the new evaluation mechanisms asked for progress reports, relating to agreed outputs and a numerical record of all activities and the number of people involved. From 2012, until the end of the programme in 2018, the same reporting mechanisms became steadily more prescriptive. Each partnership had to justify how it was working towards creating 'prosperous', 'learning' and 'healthy' communities. Still using a centralised, numerical database, local 'priority areas' were assigned, and partnerships were asked to record activity against a relevant 'performance measures', of which there were over 100. This change was a key point of contention for practitioners (Pearce, 2012) as, crucially, partnerships no longer had input into the internal evaluation framework, which dictated how progress was measured and reported, and therefore how their work was valued in an official capacity. At first glance, this process of results-based accountability was a clear, steady move from involving community partnerships directly in framing the evaluation parameters, and in doing so giving them influence over what was perceived as valuable, to excluding them through the production of externally determined outputs. This change speaks to the widespread pervasion of NPM techniques—the measures themselves constituted 'outputs' (Osborne, 2006; although it is worth noting that at the time they were often referred to as 'outcomes' by staff and officials).

In 2016, the Cabinet Secretary for Communities and Children announced his intention to end the programme based on its perceived failure to 'tackle poverty', despite its initial aim (Adamson et al., 2001); a crucial point for evaluation. Following this, the National Assembly for Wales' Equality, Local Government and Communities committee published a report entitled 'Communities First: lessons learnt', which critiqued the 'performance management' aspect of the ongoing, internal evaluation:

One of the weaknesses of the Communities First programme has been insufficient performance management. There are significant lessons to be learnt from this.... There should be a consistent level of data collated across Wales to allow programmes and approaches to be evaluated against each other and on a pan-Wales basis. (Equality, Local Government and Communities Committee Report, 2016, p. 11)

This position made the same assumption as all the externally commissioned evaluations, namely that the value of community development work *could* be measured in a uniform, metric and comparative way. For community development workers on the ground this assumption was highly questionable.

2.2 | Findings

Findings are presented here under three themes, each of which speaks to characteristics of the NPM that drive practices on the ground. Firstly, the state led evaluation was seen by practitioners as a means of (re)defining and ultimately diminishing the value of work being carried out. Secondly, evaluation was rationalised by development workers as a form of vertical accountability, thus 'distancing' between state and community by reducing communication to a one-way transmission made on the evaluator's terms. Finally, evaluation was seen as a *lost saviour* of the programme's original ethos, and a missed opportunity to preserve the vision set out at inception (Adamson et al., 2001). This final perception pertains to Lamont's (2012) call for a diversified approach to evaluation and highlights an important and potentially valuable alternative to a dominant, hierarchical evaluation framework, a subject to which we return in our concluding discussion.

2.2.1 | Defining and producing contested constructions of value

The perception of evaluation as a way of defining and diminishing value over time was a clear, recurring theme within the data, as the value of what was widely considered to be unquantifiable work was reduced to numbers. The subsequent notion of 'hidden' value, beyond the scope of official evaluation emerged later in the narratives:

but you can't put numbers on it you can't evaluate ... that woman from [X area] ... the way that she changed and her life, was more fulfilled and she knew about her options rather than just accepting things that were thrown her way, she could take charge of her own life, she got educated, she got the skills for a job, how on earth do you put a numerical figure on that? Let's just give her a twelve is it? It's meaningless. (Community development worker 3: February 2009)

And

And it was really emotional, to see these kids come in and go, right, I've gone from that step to that step. For them, that journey was as big as somebody else going from you know, gaining a GCSE to gaining a degree. Right... And at the end of it, we had a celebration for it. And they were ... they were different people... Now, how do you evaluate that? You know, these kids, suddenly from not wanting to go to school, suddenly think well we're gonna get a job. And they got a job, one of 'em got a job with um ... with a construction company. He was nowhere near ever going to get a job ... before that. (Community development worker 6: April 2017)

Participants ultimately saw the social, emotional, or relational effects as more valuable to the context in which they worked, than the number of people who had undertaken employability skills training, for example. Though both were discussed in relation to each other.

... you're measuring the wrong things for an anti-poverty programme, you know. You're measuring spend ... and, you know, you should be measuring social value ... you know. (Community development worker 7: May 2017)

Participants in 2009 and then again in 2017 consistently pointed to an incompatibility between community development work and metric evaluation, resulting in devaluation of the significance of their everyday work on paper. This devaluation was understood to obscure valuable aspects of a community development approach to tackling the effects of poverty.

Firstly, it was understood that the state valued the process of quantifying the impact of community development more than the 'actual' contribution of the work itself. Thus, the process of evaluation was seen as obscuring other forms of legitimate work, which went *underneath the wire*:

Well anyway {sighs} so yeah, that was all about um, us saying to [the monitoring body], we'll get so many people, we'll get them to with this, whatever, and we had to churn out certificates and do all that. So, we did that, but because we were experienced and knew that, that was something, but the other stuff that we supported people's lives with, was going in around that, but we never ... they never asked us about that. So we never told them. But the real support went around underneath the wire really. (Community development worker 11: May 2017)

Invisible or hidden value was also viewed in terms of the work that was having a positive influence on people's lives, but that *would not appear on any evaluation*:

when I asked some people in X area what impacts they felt the programme was having on their lives, this was after we'd been gone a couple of years and one of the women said it gave me the courage to leave my husband. That won't appear anywhere on an evaluation or it won't appear on any impact set, nothing but it's made a difference to her life, her children's lives because they'd been in an abusive, grown up in an abusive household because she had the confidence to say to him that's it, you're out and if Communities First had never come along would that have happened? I don't know you know. (Community development worker 10: October 2017)

The quote shows clear references to devaluation of certain forms of work, caused by the processes of categorisation and legitimisation produced through evaluation practices: the power to change people's lives through community development work was not measured, which in turn was positioned as a disempowering process.

Long-term impact for individuals was discussed as a valuable, yet undervalued, aspect of community development work. Often put in the context of a person's *full* journey, or a form of ongoing support beyond the parameters of a given (and measurable) event or intervention:

Say for instance, there's um, err, a young girl ... she's twenty two, I've worked with her since she was eight, she wanted to do dance, so we put dance on [for] her and her friends, and developed that, um did loads of um learning and they did the dance leaders and whatever else, now she's running her own dance school within the community for other young people to come to. You know, it's not like a, you know, we measure something over three months, six months and year but then this is like a fourteen year period. Like it's a full journey and like we've worked with that family right throughout you

know and the impact we've had is probably quite significant. (Community development worker 23: May 2017)

Long-term impact of development work for the community—tracing the *ripples* on a pond—was seen as occurring beyond the parameters of evaluation, and excluded from the short term or fixed evaluation cycles:

I think [the programme has] had a massive impact and I don't think we realise quite what impact it has had because it's like putting a pebble in a pond, the ripples just go out and out and out and I don't think we've really grasped how far over sixteen years those ripples have gone ... in some cases I don't think we'll know the impact till ten years' time. (Community development worker 10: October 2017)

The inclusion and exclusion of certain facets of the work was clearly recognised and problematised by the development workers. In these quotes, *value* was ultimately constructed in terms of an incremental, long-term and positive change for individuals and families living in poverty—*value* that was excluded from the bureaucratic records in a process of evaluation. It is worth noting, at this point, that the civil servant gatekeepers interviewed for this study were also highly critical of the evaluation. However, their focus was on the difficulty in measuring and standardising a highly variable, patchworked, community development programme in order to make it comparable within and between geographical areas. This view is in subtle, but important, contrast to the development workers' focus on the difficulties of justifying the value of community development work in the face of such attempts at measurement. Thus, conflicting notions of value were apparent from different levels of the programme.

However, many community development workers also viewed the evaluation itself as mundane, a view reflected in the recurring description of evaluation as a *tick-box process*. Alongside this, the volume, detail and emotion of discussions related to evaluation in the data demonstrated an awareness of the power embedded in the monotony of this routinised requirement. Reduction of evaluation (*to a tick box process*) in community development narratives could, therefore, be a direct response to the diminishment of community development work (*to a set of numbers*) by evaluation systems and policy narratives. A process of mutual reductionism causing frustrations on both sides.

Yeah because you can't ... measure what you've said you were going to, so the way I did it over here was I was always coming up with new things anyway, and it sort of kept them off our backs. (Community development worker 3: February 2010)

Over time, as evaluation was more driven by metrics, participants acknowledged its detrimental impact upon the day-to-day delivery of community development work and the wellbeing of the workers themselves:

You've got to really struggle in your working day to find the time to think things through, you know. And that's not good, it's not healthy, that's not, it's not healthy for your mind, you know because you're just turned into like a robot, you know, just delivering, and delivering all the time. (Community development worker 7: May 2017)

This quote showed an awareness of the significance of the selective assignment of value and its detrimental implications for those not in power. Because substantive parts of community development work were not valued enough to be included in evaluation, and therefore excluded from the process according to the parameters set out by the state (and the funder), we begin to see a systematic and long-term reinforcement of hierarchical state-community relations. We also begin to see a dual devaluation of day-to-day community development work through

the process of evaluation: both officially on paper and through its sapping effect on time and energy, which would otherwise be spent working with people.

However, the importance of flexibility and a willingness to work beyond a job description was a consistent theme throughout the data, which culminated in narratives depicting evaluation as a point of resistance. Working both within and beyond evaluation frameworks allowed community workers to be adaptable, and provide support to people when they needed it, and in the way they needed it. To listen, to be present, to go the extra mile:

So I would come into work to do something on a Monday and I end up doing it on a Wednesday or Thursday or Friday or the following week because of the amount of people coming through the door, or problems that we would sort out, that weren't ever recorded. (Community development worker 23: May 2017)

And:

They'll contact me and say "Right, I want to do this project, I've got this funding, I want to do this in your community with you, do you want to do it?", and I do it, and it may be nothing to do with our targets, but I knew the impact that that would have on our community. (Community development worker 23: May 2017)

Working outside the metric parameters set by the state was a source of pride amongst development workers, something which allowed them to do their job well and in accordance with their values. To do so they appeared to reject the encroachment of evaluation in order to properly engage with community members, and for many, in order to feel human: 'I think in the past few, couple of years we feel that we have become like ... robots' (Respondent 21, March 2017). This pertains to Certeau's notion of resistance to technocratic structures through 'tactics' which are 'articulated in the details of everyday life' (Certeau, 1984, p. xv). The question of *what is valued* by community workers versus the government formed the crux of a widening rift in hierarchical relations, accompanied by emphasis on measurable impact and particularly employability.

2.2.2 | Evaluation as a mechanism for securing forms of vertical accountability

It was also clear that community development workers felt they were being held to account through the process of evaluation, not least due to the one-way flow of information. Outputs submitted as part of the evaluation process, were sent 'upwards' from local partnerships to local authorities, from local authorities to Welsh Government, civil servants and from there to Ministers. It was viewed by many as a unidirectional form of information conveyance because numbers were submitted without clear explanation around what the information was being used for. In addition, due to the exclusion of large, rich parts of the work, it was perceived as narrow and prescribed—a restricted or simplified avenue through which much more complex messages failed to be conveyed:

they say, "still send us the case studies", the difficulty is ... we don't hear anything back. (Community development worker 7, May 2017)

However, this conveyance was also viewed as a way of keeping community groups at arm's length and maintaining symbolic distance, using the formal evaluation structures:

And they send people in who are completely and utterly separated so much from the people that they're working with. And they might do a little bit of work and they might be really good, but then

they just go. They just like, they just disappear. And that, you know that's not good for people who have got so involved in a project. You know, they promised this that and what have you. And then they completely, they just disappear. They've got their bit, they've written up their results. Done their evaluations, got paid nicely for it and disappear. And I think that's just leaving people to fail. (Community development worker 19: February 2017)

This type of distancing between policy implementation and policy makers is a key characteristic of the NPM (Osborne, 2006). This was also the most direct or explicit form of perceived hierarchical reinforcement, highlighted by the word 'upwards', used when referring to reporting to local authorities to civil servants and then to Ministers. The format, frequency and use of this information was entirely dictated by the Government from 'above', its enforcement by the local authority and its production by the community partnership. This division of labour clearly belied a hierarchical order of governance.

2.2.3 | Evaluation as a 'lost saviour'

Going further, evaluation was seen as a lost saviour of the programme's original community development ethos (Adamson et al., 2001) and based on early Welsh Labour ideologies (Davies & Williams, 2009; Dicks, 2014; Pearce et al., 2020). Numerical measurement of value excluded substantive parts of community development work, and evaluation came to be seen by multiple respondents as a lost defender of a practice being held to account:

... in those days it wasn't, we weren't so clever about outcomes and what had we achieved ... if only I had had the benefit of hindsight I would've made sure that we documented things better ... an awful lot of money was invested in those communities and people are bound to ask well what difference does it make and I think that's the mistake I might've made ... and I think if we'd put more effort into finding out the impact of what we were doing and measuring that impact we would've been in a much better position when people wanted to make changes to the programme.... (Community development worker 10: October 2017)

Here, the respondent was taking responsibility for what she saw as an inadequate degree of localised evaluation by community development workers in the early days of the programme; evaluation was positioned as a tool with the potential for challenging the state's monopoly over the definition of value. Crucial to this, participants differentiated between *the* evaluation imposed (by the state) and the responsive and reflexive evaluation that could have been carried out by development workers themselves. The latter increased a sense of professional worth, reward and improved working practices:

The boy from the house, and the morning we picked him up to go right he had been sprayed with tear gas because there was a raid on the house. And as we're going round he was in awe of the ceiling in the cathedral. So at this residential we done, we went on a boat trip, we done go-karting. All the things young people like to do, all really good things. And on the [day trip] evaluation you know what he enjoyed best? The cathedral. The cathedral. And this thing, people say oh when you take people, young people bowling. That's the whole idea of the evaluations.... (Community development worker 12: March 2017)

In this example, the community development worker designed and administered the evaluation herself. She was still accountable, her work was still being evaluated, but the answer to 'knowledge for what and for whom?' was very different, derived from her experience of her community development work. The knowledge gained was positioned

as authentic and legitimate, used to improve working practices for those providing the service, rather than to hold them to account. In this case information was transferred horizontally, in keeping with Lamont's (2012) notion of *heterarchies*. Such constructions acknowledge the versatility of evaluation in terms of meaning and interpretation, which in turn was related to the intended purpose. The value of evaluation was largely undisputed, but the question of value 'for whom and for what' remained a point of contention.

This view of evaluation—as a tool for change—makes the assumption that political decisions draw directly upon evidence in a linear, rational approach to policy making; and that the purpose of evaluation is to collect and use knowledge. However, in cases where, what was considered, more authentic and legitimate forms of evaluation had taken place, the perception on the ground was that this was not seen to be a persuasive form of evidence. In the following quote, for example, the development worker discussed a project that was evaluated using action research methods. It was framed as collaborative, substantive, capturing the full range of community development activity taking place, using mixed methods and resulting in clear, actionable findings which were fed back to practitioners and policy makers. However, it was felt that this robust body of research had no impact on the government's decision to cut funding and end the project.

They didn't follow our model at all. So even though we proved that having a support worker in addition to a tutor, for people who really lack confidence in learning, and have ups and downs and complex lives ... they went back to ... to delivering adult ed ... because that's all they knew. Which is ... it is a crying shame, absolutely crying shame. And my thing is if that could happen there, how many other places has that happened? (Community development worker 11: March 2017)

The perception of evaluation as a lost saviour was therefore based on the assumption, that evaluation of this type should be used by policy makers to improve their programmes, and that less output focused data showing the full effect of community development work should have protected worthwhile projects. However, the concept of a lost saviour more strongly re-emphasised the perceived power of metric-driven evaluation over day to day working practices and perceptions of self-worth, while having little impact in terms of policy-driven or state-led change. In short, evaluation is shown here to reinforce perceptions of hierarchical structures while having limited implications for policy decision-making.

3 | CONCLUDING DISCUSSION

The reinforcement of hierarchy through state-led evaluation is made visible through the accounts presented here, which reflect upon and grapple with the wider significance of an everyday evaluation practice. Our analysis identifies three ways in which evaluation processes can reinforce (and resist) hierarchical relations between state and civil society, which further aid the development of evaluation as a concept. Firstly, as a means of defining and ultimately producing contested constructions of value. Hierarchical relations are reinforced through the exclusions of certain types of practices, whereby certain types of value are hidden. Respondents are aware of the delineation of value through the omission of substantive elements of their work, and that such omissions prioritise metric measurement over richer conceptions of social value. Secondly, evaluation is positioned as a mechanism for securing forms of vertical accountability; notions of hierarchy are further enforced by the distancing of policy makers from policy implementation. Community development worker narratives consistently present the metric information they were required to collate and submit as 'useless' in terms of improving their work, partly because of the auditory form it took and partly because it was not an instigator of wider change. Crucially, here we see an assessment of the purpose of evaluation. In terms of asking *knowledge for whom?*, if the data was not for those producing it, then it must be for those who requested it and in terms of *knowledge for what?* if the findings have no actions associated with

them, then the purpose is unclear. In the absence of clarity, the process of evaluation was positioned as a form of distancing and holding-to-account, both characteristic of NPM. Finally, through its construction as a lost saviour, evaluation as an entity is seen to have untapped potential for safeguarding the integrity of professional and political ideology, while also highlighting the ineffectiveness of the process in guiding policy or changing government actions.

Significantly, there is little direct discussion around structural issues causing change to the programme; no mention is made of wider formations of neo-liberalism, the austerity agenda, welfare reform or the ruling governments—all key to the eventual demise of the programme. The primary way in which participants engage intellectually with the social problems and pressures associated with the state is via their critique of *the* evaluation. It is their intuitive sense of annoyance at grappling with these processes that makes the respondents cognisant of the wider power structures and agendas in which their work is enmeshed, and thus able to preserve elements of their ideological principles in practice. Complicity with and contestation of evaluation are both visible in the data, in ideas of 'ticking the boxes', 'keeping them off our backs' and doing the valuable work 'underneath the wire'. Both are acknowledged and practiced. Founding principles of empowerment and capacity building are maintained in the face of tightening directive, and while the spaces to practice them contract, they do not disappear (Blakely, 2011). However, it is important to note that the role of protector against encroachment by mechanisms of accountability, consolidated by phrases such as 'I keep them off our back', is often observed at the expense of the worker's time and mental health. The work perceived as most valuable operates outside the official categorisation and legitimisation processes, and precisely because it is excluded, it is not held to account by a set of opposing principles. The value of work that went 'underneath the wire' presents a very different picture of the programme's perceived 'failure to tackle poverty' which was the basis for its termination.

Here, we see a privileging of certain types of knowledge over others within social policy evaluation, which reinforces hierarchy while also providing a point of understanding from which it is possible to grapple with and reflect upon more abstract issues of austerity, reform, recession, labour market precarity and welfare. We believe these insights can potentially form the beginnings of a higher degree of abstraction in which to treat evaluation for future analyses. With respect to recommendations for policies and practices, the data and analysis presented here indicate the need to consider practitioner knowledge and experience beyond metric inputs in future. Following Bartels (2017, p. 3801), we argue that social innovation is typically 'contested, challenged and resisted by institutionalised ways of working' (Bartels, 2017, p. 3801). Strong and democratic communication channels embedded in civil society are needed if lessons are to be learned from innovations and policy experiments. The hegemonic institutions of the state, including evaluation, require considerable reform if they are not to marginalise the role of civil society. In the case of Communities First an evaluation fit for a community development programme would include a more collaborative and qualitative element. Indeed, this was an integral part of the programme's ethos, at least for some of those involved in creating it (Adamson et al., 2001). Such arguments for strategic level collaboration and partnership are visible in wider literature on democratic experimentalism and social cohesion (Moss, 2012, p. 10), but manifest most specifically in Realist(ic) approaches to evaluation. Drawing on the Realist(ic) approach 'to lead better-focused and more effective programmes' (Pawson & Tilley, 2004, p. 15), co-production and collaborative evaluation would be a natural match for a programme rooted in empowerment and capacity building. The Realist(ic) research cycle begins with a co-produced hypothesis, then data collection and testing the hypotheses using the data; finally, assessment of the analysis. All four stages 'take an insider perspective' whereby the knowledge of stakeholders is paramount to developing a *shared* understanding of the programme in question (ibid, 2004, p. 12). Such a shared understanding is absent from the narratives presented by practitioners. In contrast, a Realist(ic) approach, whereby the evaluation process can provide a locus of shared external control, could create heterarchy to counterbalance political, social and economic hierarchies (Lamont, 2012). Using Pawson and Tilley's (2004) framework, this could be achieved through the meaningful inclusion and

representation of practitioners in early decision-making and the co-development of hypotheses to interrogate the data collected through a range of sources, including at policy level (Marisol and Gaventa 1998). Future evaluations of this type could therefore be made more effective through co-production and collaborative evaluation in its widest and most inclusive sense.

ACKNOWLEDGEMENTS

Thanks are due to the Economic and Social Research Council for funding this project (ES/L009099/1). The project was also partly funded by the Arts and Humanities Research Council who support Productive Margins—a collaboration between Cardiff and Bristol University—who we thank for supporting data collection—and in particular Dr. Eva Elliott of Cardiff University, Sue Cohen of Bristol University and Allan Herbert of the South Riverside Community Development Centre. Thanks are also due to the Wales Centre for Social and Economic Research and Data (WISERD) for supporting the fieldwork. Finally, thanks to Professor Sally Power and Dr. Jamie Lewis for their valuable comments and suggestions on the paper.

ORCID

Sioned Pearce  <https://orcid.org/0000-0001-6906-1096>

ENDNOTES

¹ Political ideology in this context refers to the wider approach to policy undertaken by Welsh Labour in relation to Communities First.

² Communities First areas averaged between 1,500 and 11,000 in population and were at ward, sub-ward and lower super output level.

REFERENCES

- Adamson, D., & Bromilley, R. (2008). *Empowering communities in practice: The case of communities first in Wales*. Joseph Rowntree Foundation: York.
- Adamson, D., Dearden, H., & Castle, B. (2001). *Community regeneration. Best practice review*. Cardiff: National Assembly For Wales.
- Allen, K., Needham, C., Hall, K., & Tanner, D. (2018). Participatory research meets validated outcome measures: Tensions in the co-production of social care evaluation. *Social Policy and Administration*, 53(2), 311–325.
- AMION Consulting Limited, & Old Bell 3 Limited (2011). *The evaluation of Communities First*. Cardiff: Welsh Government.
- Bartels, K. (2017). 2017 The double bind of social innovation: Relational dynamics of change and resistance in neighbourhood governance. *Urban Studies*, 54(16), 3789–3805.
- Bevir, M., Needham, C., & Waring, J. (2019). Inside co-production: Ruling, resistance, and practice. *Social Policy and Administration*, 53(2), 197–202. <https://doi.org/10.1111/spol.12483>
- Blakely, H. (2011). 'A second chance at life': Labour, love and welfare on a South Wales estate. PhD Thesis. Cardiff University.
- Bovaird, T. (2012). Attributing outcomes to social policy interventions—'Gold standard' or 'Fool's gold' in public policy and management? *Social Policy and Administration*, 48, 1.
- Burawoy, M. (2004). For public sociology. Presidential address at the American Sociological Association.
- Burawoy, M. (2012). The public university—A battleground for real utopias. Presentation at the annual meeting of the American Sociology Association.
- Cambridge Policy Consultants. (2006). *Interim evaluation of communities first*. Final Report.
- Certeau, M. (1984). *The practice of everyday life*. Berkeley: University of California Press.
- Coleman, J. (2009). *Communities first*. Welsh Audit Office: Cardiff.
- Davies, N., & Williams, D. (2009). *Clear Red Water: Welsh devolution and socialist politics*. London: Francis Boutle Publishers.
- Dicks, B. (2014). Participatory community regeneration: A discussion of risks, accountability and crisis in devolved Wales. *Urban Studies*, 51(5), 959–977.
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113, 1–40.
- Foucault, M. (1980). *Power/knowledge: Selected interview and other writings. 1972–1977*. Harlow: Prentice Hall.
- Freeman, R., & Maybin, J. (2011). Documents, practices and policy, evidence and policy. *Journal of Research, Debate and Practice*, 7(2), 155–170.

- Gove, M. (2020). The privilege of public service, the Ditchley Annual Lecture. Gov.uk. Retrieved from <https://www.gov.uk/government/speeches/the-privilege-of-public-service-given-as-the-ditchley-annual-lecture>
- Hincks, S., & Robson, B. (2010). *Regenerating communities first neighbourhoods in Wales*. Joseph Rowntree Foundation: York.
- Ipsos MORI and Wavehill Consulting. (2015). *Communities first: Process evaluation*. Welsh Government.
- Lamont, M. (2017). *Prisms of inequality: Moral boundaries, exclusion, and academic evaluation*. *Praemium Erasmianum Essay 2017*. Amsterdam: Praemium Erasmianum Foundation.
- Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38, 201–221.
- Lamont, M., Beljean, S., & Clair, M. (2014). What is missing? Cultural processes and causal pathways to inequality. *Socio-Economic Review*, 12, 573–608.
- Larner, W. J., & Butler, M. (2005). Governmentalities of local partnerships: The rise of a 'partnering state' in New Zealand. *Studies in Political Economy*, 75, 85–108.
- Lowe, T., & Wilson, R. (2015). Playing the game of outcomes-based performance management. Is gamesmanship inevitable? Evidence from theory and practice. *Social Policy and Administration*, 51(7), 981–1001.
- Marisol, E., & Gaventa, J. (1998). *Who counts reality?: Participatory monitoring and evaluation: A literature review*. Institute of Development Studies: Brighton.
- Moss, P. (2012). There are alternatives! Markets and democratic experimentalism in early childhood education and care. Working Paper No. 53 Bernard van Leer Foundation and Bertelsmann Stiftung.
- National Assembly for Wales. (2003). Social justice and regeneration committee paper (SJR-08-03(p.2)). *Communities first review*.
- National Assembly for Wales. (2017). *Communities first: Lessons learnt*. NAFW, Cardiff.
- Newman, J. (2005). *Remaking governance: Peoples, politics and the public sphere*. Bristol: Policy Press.
- Osborne, S. P. (2006). The new public governance? *Public Management Review*, 8(3), 377–387.
- Parr, S., & Churchill, H. (2019). The troubled families Programme: Learning about policy impact through realist case study research. *Social Policy and Administration*. Early View.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation use*. London: The Guilford Press.
- Pawson, R., & Tilley, N. (2004). *Realist evaluation*. London: British Cabinet Office.
- Pearce, S. (2012). *Constitutional change and community development: Communities first under the Welsh government*. Sheffield: The Centre for Regional, Economic and Social Research.
- Pearce, S., Sophocleous, C., Blakely, H., & Elliott, E. (2020). Moments of alignment between devolved political ideology and policy design: The case of Wales. *People, Place and Policy*, 14(1), 6–23.
- Peters, B. G. (2018). *Policy problems and policy design*. Cheltenham: Edward Elgar Publishing.
- Pollitt, C. (1995). *Justification by work of by faith? Evaluating the new public management*. *Evaluation*. New Delhi: Sage.
- Power, M. (1997). *The audit society: Rituals of verification*. London: Oxford University Press.
- Prior, L. (2008). Repositioning documents in social research. *Sociology*, 42(5), 821–836.
- Richardson, L. (2013). Putting the research boot on the policymakers' foot: Can participatory approaches change the relationship between policymakers and evaluation? *Social Policy and Administration*, 47, 4.
- Roethlisberger, J., & Dickson, W. J. (1941). *Management and the worker*. Cambridge: Harvard University Press.
- Sabel, C., & Simon, W. (2017). Democratic experimentalism. In J. Desautels-Stein & C. Tomlins (Eds.), *Searching for contemporary legal thought* (pp. 477–498). Cambridge: Cambridge University Press.
- Scorer, J., & Adamson, D. (2007). *Communities first in RCT: Community focus groups and professional panels baseline report*. Pontypridd: University of Glamorgan.
- Timmermans, S., & Epstein, S. (2010). A world of standards but not a standard world: Toward a sociology of standards and standardization. *Annual Review of Sociology*, 36, 69–89.
- Vedung, E. (2017). *Public policy and program evaluation*. New York: Routledge.
- Wavehill Consulting (2007). *Evaluation of the communities first trust fund*. Welsh Government: Cardiff.
- Welsh Government. (2006). *Communities first: A 2001 baseline*.
- Welsh Government Equality Local Government and Communities Committee, 21st. (2016). Carl Sargeant's justification to end communities first. Retrieved from <http://senedd.assembly.wales/documents/s65351/21%20June%202017.html> CT=2

How to cite this article: Pearce S, Blakely H, Frayne D, Rees Jones I. Evaluation in reinforcing and resisting hierarchical relations between state and civil society. *Soc Policy Adm*. 2020;1–15. <https://doi.org/10.1111/spol.12668>