

Vivienne Cree, Sumeet Jain & David Peter Hillen (2018) Evaluating effectiveness in social work: sharing dilemmas in practice, *European Journal of Social Work*, DOI: [10.1080/13691457.2018.1441136](https://doi.org/10.1080/13691457.2018.1441136)

Evaluating Effectiveness in Social Work: Sharing Dilemmas in Practice

Abstract

Evaluating effectiveness is of primary concern to social work practice; resources will always be limited, and those using services deserve the best. But what is effectiveness, and how should it be measured? This paper explores the challenge of measuring effectiveness in social work by examining two small-scale evaluations of services (statutory and voluntary) in Scotland. Both evaluations used a mixed method approach and both were conducted in a climate of funding cuts. It will be argued that evaluating effectiveness sits uncomfortably between new public management (NPM)'s imperative for measurement and efficiency within a market economy and agencies' own need to reflect on their practice and respond to the views of their service-users. Such processes force agencies into a continual cycle of monitoring and review that may, paradoxically, impede organic change and development. It is suggested that an alternative, critical paradigm in evaluation offers a positive way forward. (147)

Keywords: evaluation, new public management, social work, critical paradigm, research methods, realistic evaluation

Evaluating Effectiveness in Social Work: Sharing Dilemmas in Practice

Introduction

Evaluating effectiveness in the delivery and management of social work services is a wholly desirable objective. It is essential that service users get the opportunity to report on the services they are receiving; it is important too that those who are paying for and providing services assess the degree to which services are meeting identified needs (or not) and whether change is necessary to bring about improvements in the quality of care or quality of life for service users. It matters little who is paying for or providing services: statutory (public sector), voluntary (third sector) or independent (private sector, both 'for-profit' and 'not-for-profit') agency. In each situation, evaluation is a crucial component. It is what keeps agencies and services alert and responsive and is what helps guard against complacency and atrophy; as Shaw (2011: 9) argues, social work and those who use our services *need* social workers who understand and are committed to evaluating their practice. Evaluation is also hugely important for the public image of social work. Social work and social workers have consistently had to confront a hostile press and unsympathetic politicians. They are traditionally 'damned if they do and damned if they don't', blamed for doing too much and, at the same time, criticised for not doing enough (Ayre, 2001).

Evaluation is therefore not only necessary, but is also central to social work's value-position. But evaluating effectiveness in social work has some innate challenges. As we will describe, reforms under the broad heading of new public management (NPM) (Gruening, 2001; Rogowski, 2010) initiated by UK governments from the late 1970's up to the present day have pushed evaluation in particular directions, leading, we will suggest, to increased pressure on agencies to demonstrate performance, at times, at the expense of effectiveness. This leads us to call, as others have done already (see Bamberger *et al.*, 2012; Cross *et al.*, 2010; Pawson and Tilley, 1997; Shaw, 2011) for

a different approach to evaluation, one which understands the political context, and, at the same time, is honest about its claims while seeking to co-produce assessments in ways that are meaningful and relevant to those who use and deliver services.

We firstly locate social work evaluation in its historical and political context. We then present two recent evaluations: one conducted in the statutory (public) sector and the other in a third sector (voluntary) agency, demonstrating that both statutory and third sector agencies feel the pressures of NPM. We conclude by exploring more broadly key issues that small-scale, service evaluations such as these raise, arguing that it can be difficult for agencies to manage the competing demands of NPM's bureaucratic imperatives with their own need for self-reflection and improvement.

Evaluating social work: a historical overview

There is nothing new about acknowledging the need to evaluate social work practice. On the contrary, we need look no further than the social work pioneers of the early twentieth century and the work of Mary Richmond (1913 and 1917) to find acceptance of the idea that we must think critically about what we are doing in social work and why. Building on her extensive study of Charity Organization Society case-records, Richmond developed the methodology of 'social diagnosis', in which she asserted that care had to focus on the person within their situation; a focus on one without the other would not, she asserted, bring about positive change. But it was in the 1970's that the significance of evaluation was understood fully for the first time. In 1970, two British social work academics published an account of casework practice in the Family Welfare Association from the points of view of service users (then called 'clients') and practitioners. *The Client Speaks* (Mayer and Timms, 1970) became a classic text, in which attention was drawn to the disparity between client and social worker realities. Mayer and Timms concluded that to be effective, social

workers must attend to both emotional and practical needs; they must also show competence and authority in their work. A few years later, in the US, Fischer (1978) argued that 'the issue of the effectiveness of practice must always be of paramount concern to the profession' (p 8). His appraisal of the research evidence for social casework as a method was damning. Reviewing research studies from the 1930's to the 1970's, he found that social casework had no measurable impact on the lives of clients. Worse still, he concluded that 'professional casework may be hazardous to our clients' well-being' (1978: 109). Fischer's work was followed by other UK-based research on the impact of social work practice, including Rees and Wallace (1982) and Lishman (1984). While this research was less overtly negative about social work practice, it nevertheless identified a mismatch between the expectations of service users, practitioners and the general public.

Although the motivation behind these studies was profession-focused, the context within which they took place was without doubt the neoliberal revolution that was beginning to sweep across both the UK and the US, as Margaret Thatcher and Ronald Reagan's governments sought to cut back on welfare spending and reduce the influence of the public sector and public sector workers (Harvey, 2005). Ferguson and Woodward (2009) point out that at this time, those on the Left and Right of the political spectrum in the UK were highlighting similar concerns about the need to restructure public services and weaken professional power. The proposed solutions, however, were vastly different: greater community initiatives and stronger trade union and user movements were proposed by the Left; the introduction of a market of care by the Right. Meanwhile, 'the erosion of the welfare state and the privatisation of public services began' (Ferguson and Woodward, 2009: 38), just as managerial terms such as 'value for money' and 'consumer' gained prominence, and key social work services were 'contracted out' to third sector and private agencies (ANON, 2009). The 1990's saw the publication of a raft of books and reports promoting

evaluation, demonstrated in the work of Cheetham *et al.* (1992), MacDonald and Sheldon (1992), Barnes (1994), Everitt and Hardiker (1996), Shaw and Shaw (1997), Cheetham and Mansoor (1998), Cheetham *et al.* (1998), Trinder (1996 and 2000), Fisher (1999) and Gomm *et al.* (2000).

The new attention to evaluation was different from before. Whilst it had always been thought important for the profession of social work to give attention to what it was doing and why, much of the impetus for evaluating services from the 1990's onwards came not from the profession, but from government, and specifically in the UK, via new public management (NPM). Banks (2011) argues that although NPM arrived with the 1979-1997 Conservative government, it was with New Labour between 1997 and 2019 that it really intensified, with the expressed goal of 'modernizing public services'. What was seen at this time was an increasing diversification of provider markets in social work and social care, with local authorities being encouraged to purchase services from independent 'for-profit' and 'not-for-profit' agencies. Alongside this, ever-stronger 'quality measures' were required by purchasers 'to aid them in their commissioning decisions and to contribute to the monitoring of aspects of performance, such as efficiency and effectiveness' (Malley and Fernández, 2010: 559). As Ferguson and Woodward assert:

' A commitment to managerial reforms, based on notions of 'what works' and 'best value', and driven by the discipline of the 'three E's' – economy, efficiency and effectiveness – was at the heart of New Labour's programme' (2009: 42).

Evaluation became, in effect, a central constituent in NPM, a key part of the system that regulates and disciplines social work practice and social workers themselves. The fiscal crisis of 2008, the 2010 Coalition government and the Conservative government since 2015 led to yet more 'welfare reform'; the so-called 'austerity'

agenda has steered further public sector cuts, as increasingly services are contracted out or sold to third sector and private sector providers. It is argued that if public services are to continue in local authorities, they must demonstrate further efficiency and effectiveness; meanwhile, 'the demand for performance measures, including value for money and social return on investment, is as important as ever' for both third and private sector agencies (Banks, 2011: 11).

This, then, is the backcloth against which the two evaluations were set. Neither is offered as an example of 'best practice' in evaluation, should such a thing exist. Nor do we discuss the findings of the evaluation; these have already been shared with the agencies and executive summaries were sent to all those who took part. Instead, building on previous publications (ANON *et al.*, 2016; ANON *et al.*, 2014), we use our experience to help us unpack and connect with an alternative approach to evaluation that foregrounds a critical methodology.

It should be noted that in both situations, the evaluation proposals and methods were scrutinised and approved by the researchers' host university's ethics approval process.

Two case study examples

The statutory sector project evaluation

The first evaluation was conducted in 2012 by a team of three researchers from a university in Scotland. A local authority commissioned this evaluation of two of its services, one focused on people with alcohol-related problems and the other, on people with drug-related problems. Both services received funding from a mix of local authority and local health board (funded by central government) through a local alcohol and drug partnership (ADP), and both provided a dedicated social work service to clients.

Alcohol and drug misuse are key priorities for the Scottish Government. There had been a strong policy drive in this area, and a marked shift in direction from 'harm reduction' to 'recovery', with the added identification that substance misuse harms families and communities as well as individuals (Scottish Government 2008, 2009a). The Scottish Government had given guidance to ADPs about what and how services should be provided. The *Delivering Better Outcomes* 'toolkit' published in 2009 heralded the move to an 'outcomes approach', with national outcomes, targets, and indicators specified, as a means to 'improving service delivery and recovery' (Scottish Government, 2009b). The toolkit stated that ADPs must seek to make connections between service delivery outcomes, intermediate outcomes, high level outcomes and national outcomes; the purpose of this evaluation was to assess how far the services were meeting their specific outcome measures. Some of these measures were similar across the two services but others were different. The alcohol service's outcome measures were: 'reduction in alcohol consumption; improvements in employment and accommodation status; improvements in financial situation; improvements in contact with non-alcohol users; and more and better use of a range of services'. The drug service's outcome measures were: 'reduced use of (illicit) drugs; improved psychological health, reduced chaos in daily living; increased self-worth; increased access to a wide range of treatment and other therapeutic interventions; children being safer'.

The parameters for the evaluation were set by the services' manager and team leaders in advance of the evaluation proper. It was specified that there was to be no specific examination of inputs (financial resources e.g. staffing, resources etc.), and instead, the focus would be on outputs (activities, programmes) and outcomes (short and longer term) with a view to making a judgement about the effectiveness of the services (assessing what progress had been made to meet the outcome measures).

A mixed method approach was used (Wilkinson, 2004), including both qualitative and quantitative methods, and research ethics permission was sought from and granted by the authors' host university. We firstly met staff to plan the evaluation then conducted a targeted literature review of relevant national and local policy documents and research evidence in order to scope the field and understand more fully the broader strategic outcomes that the services were seeking to address (Boote and Beile, 2005; Hart, 1998). This was followed by an analysis of services' reports and case-records (looking at service user demographics, care plans, case-notes, reasons for closure, feedback from service users etc.) and participant observation (shadowing social workers on visits to service users). Focus groups were conducted with social work practitioners from both services and then interviews were undertaken with different stakeholders (the services' manager, team leaders and a sample of 20 referrers and 20 service users). The evaluation report was shared in draft form with the manager and other agency staff before finalising. The team then gave an input to senior council staff.

We now turn to the voluntary sector service evaluation.

The voluntary sector project evaluation

The second evaluation took place in 2015, and was carried out by two of the researchers who had worked on the first evaluation. On this occasion, the evaluation focused on a project that offers one-to-one and/or group-work support to children who are living in a Scottish city in a family where there has been domestic abuse. The project is one of a number of services run by a voluntary sector social work agency; many, but not all, of the families that use the agency are from black and minority ethnic (BME) communities and some are new migrants to Scotland. The agency relies on funding from a mixture of sources: local government, private agencies and fund-raising, and each of its services comes up for funding renewal at

regular intervals. The evaluation was thus a necessary part of securing funding over the next three years; like the statutory evaluation, the focus was to be on measuring how far the service's outcome targets had been met since the last evaluation had been carried out three years before. Significantly, the project was working to two different sets of outcome targets for its two charitable funders. The first detailed these as follows: 'young survivors feel that support services have helped them to cope and improve their mental and emotional well-being; young survivors feel that support services have helped to keep them safe and increase their resilience against future harm; young survivors feel that having a voice and hearing from fellow survivors has helped them to progress their healing process'. Only the first two were listed in the second funder's 'Outcomes and Indicators' Review' proforma.

Again, the limits of the evaluation were clear from the outset. There was to be no access to staffing or financial records, and a very tight timescale and small budget meant that what could be achieved was inevitably constrained by both time and money. In common with the statutory evaluation, this one used a mixed method approach. At the outset, research ethics permission was sought from and granted by the authors' host university. An early exploratory meeting was held with staff to plan the evaluation and a brief literature review was undertaken to consider the context of the project's work more fully. A targeted analysis of services' reports and case-records was undertaken. Ethnographic observation of a group session with children was conducted, followed by a focus group discussion with 5 children. Finally, interviews were carried out with different stakeholders (3 project staff, 3 referrers and 2 parents). The evaluation report was shared in draft form and discussed at a meeting with all agency staff before finalising.

Shared issues

Although the evaluations were, in many ways, different (different sector, different service-user group, different funder), they shared common features which, we believe, are common across many small-scale, service evaluations that are conducted in social work and health settings today. Most importantly, neither looked at inputs, and both focused instead on attempting to identify how well each service had met its stated outputs and outcomes. There were difficulties and tensions in all aspects of this - inputs, outputs and outcomes - as we now discuss.

Inputs

Because we did not have access to inputs (for example, internal financial documents or staff figures and appraisals), we were unable to reach any formal assessments of cost-benefit, value for money or staff performance. This is not unusual, in our experience; agencies rarely have the money or resources to pay for a broader evaluation, and, we would argue, this necessarily has implications for what can be claimed in an evaluation.

But there is a seeming contradiction here. Just because we did not have access to data on inputs, this does not imply that inputs were absent either from the agenda of those whom we interviewed, or from our evaluation reports. On the contrary, staff, referrers and service users all told us a lot about the quality of the services and about the relationships that existed between staff and service users. So, for example, we learned that social workers in the statutory services wasted countless hours inputting data into what they saw as clunky, non 'user-friendly' computer databases (this issue emerges in previous research, including White *et al.*, 2010). Meanwhile, we realised that those in the voluntary agency were awaiting the arrival of a new client database which, they believed, might be the answer to all their problems with uneven record-keeping and files-management. So inputs featured in the evaluations, in spite of the fact that they were not part of our formal data collection. This chimes with the

approach proffered by Pawson and Tilley (1997), who assert that good evaluation must be real, realist and realistic; it needs to be scientific, but at the end of the day, it is *applied* research (p xiii), pursued to inform policy and practice, not build theory. From our point of view, that meant taking all available evidence into account, wherever and whenever it emerged; it meant being pragmatic about what we were finding, while, at the same time, forefronting the importance of reflecting the issues that were important to those whom we were interviewing.

Outputs

With no access to 'hard' data on inputs, both evaluations sought to find out what the outputs were and hence ascertain how far the outputs lived up to the promises detailed in mission statements and agency objectives. This proved difficult, for a number of different, but related, reasons.

Unreliability of figures – what is counted, how and why?

Both the statutory and voluntary agencies had targets for numbers of service users using their services over a given period of time, but figures can be notoriously misleading (Best, 2001). For example, the number of individuals using a service in a week and the cumulative total of the number of people going through the doors each day may be vastly different, especially when one individual may attend on numerous occasions over the course of a week. Which figure accurately records project usage? Furthermore, the way an agency deals with re-referrals and the whole process of opening and closing cases also impacts on number-counting. So, for example, the statutory sector agency only held cases open for 16 weeks; the goal over this period was to move people onto other mainstream provision. But for many service users, 16 weeks was not long enough, for a variety of reasons, including time spent in hospital or in prison, so they were re-absorbed into the service for another 16 weeks' spell. Should they, then, have been treated as 'new' cases or not? From agency records, it

was clear that the work that was involved in supporting them was the same either way.

The voluntary sector agency was also faced with stringent limitations on its services, especially on its group-work programme. Funders expected the agency to provide three sets of group sessions each year, involving different children each time. But the reality was that many children returned to subsequent groups; they wanted to, and the agency felt that they needed to. So which figures should be used? The answer to this question had a clear impact on the agency's recording systems as well as any evaluation of a single service within the overall project. There is, moreover, another issue about numbers. If an individual who is referred decides not to access a service, for a range of reasons, or attends once and not again, are they to be regarded as agency 'failures'? On a number of occasions in the alcohol service, people died while being helped. Again, are they 'failures' or not? Surely they tell us more about the age, ill-health and gravity of the problems the service is coping with, rather than about shortcomings on the part of the service? This also suggests that we may, in fact, be counting the wrong things in the world of NPM.

What these illustrations show is that counting heads is difficult in both statutory and voluntary agencies. Moreover, there is a potentially perverse incentive on agencies to massage, if not lie, about their figures to secure continued funding. We are not, of course, suggesting that either of the agencies that we were evaluating did this; rather, that numbers can be used in different ways and numbers have effects; they determine how resources (staff and money) are allocated, and hence who gets (and who does not get) a service. In this process, those who get left out are often those on the margins of the boundaries and borders as determined by NPM; they do not fit the mission statement or profile as determined by the funder. The stakes are very high in this game, as shown in the much-documented collapse of the children's charity, Kids

Company, in August 2015 (see

<http://www.theguardian.com/commentisfree/2015/oct/16/kids-company-camila-batmanghelidjh-charity>).

Unreliability of agency records – what is recorded, how and why?

It would be easy to be critical about the record-keeping systems in both agencies. What we saw in both was the variability of record-keeping, both online and on paper. In the statutory evaluation, we ended up with vast amounts of different kinds of data to analyse, and it was difficult to make sense of these. In the voluntary agency, there was a lack of tie-up between individual case records and recording of group sessions, and lots of missing referral sheets. It was evident in both agencies that social workers may record the reasons for referral and any decisions taken, but not have time to record either the process or the outcomes, unless forced to by the demands of a statutory assessment. Problems with data in the statutory setting were partly a result of the different electronic recording formats that social workers were required to use. However, over and above this, practitioners had different styles of writing and recording, with some adopting a more narrative approach than others. There were also issues with the voluntary agency's record-keeping systems, making it difficult to chart a child's progress through the agency, from point of referral to ending, with a clear picture of everything that had happened in between.

Variability in recording makes any analysis and evaluation of agency records enormously troublesome, as Hayes and Devaney's earlier (2004) work on the analysis of case files illustrates. What is an evaluation to say except that this variability exists; that there is a large amount of reflective data on the one hand and the complete absence of it on the other. What does either tell us about the effectiveness of a service? Is it possible that this tells us more about efficiency rather than efficacy?

Outcomes

If it was difficult to measure outputs, it proved just as challenging to be certain about outcomes, again for a number of related reasons.

Who determines outcomes and what should they be?

One of the most pressing issues facing agencies is how to align the hopes and aspirations of the various stakeholder groups, namely, the agencies, the funders and the service users and their families. Funders are likely to have a particular social issue that they wish to fund. Agencies may, in turn, 'tweak' their stated aims and objectives in order to secure funding, making it difficult to align objectives and outcomes in an evaluation. In our evaluation of the alcohol service, we found that one-third of the service users did not want help with their alcohol use; the support they wanted was with housing, money and family relationships, not alcohol. So what does the agency do? Decide not to offer a service to them? Or perhaps hope that their alcohol use will be helped indirectly? It is, after all, known that when people's lives are more settled, they are less likely to turn to excessive alcohol use (Galvani, 2012). This suggests that 'softer' outcome measures (e.g. willingness to work with a social worker) are likely to be as important as 'harder' ones (e.g. a reduction in alcohol use). But will the 'softer' outcomes be 'hard' enough to secure continued funding?

There is, of course, a whole industry now surrounding the idea of 'soft outcomes', with private sector consultancies and trainers offering advice on how to set up and measure these. But none of the advice that is available takes away from the reality that an individual's motives in accessing a service may be very different to that of the agency – or the funder – and that the social, economic and political context inevitably determines what is provided.

How open can a service be about its outcomes?

Social work services are most often offered to marginalised individuals and communities; services can themselves be both stigmatised and stigmatising (ANON, 2014). This has a major impact on service delivery, and hence on the evaluation of services. In the voluntary sector agency, a decision had been taken not to speak directly with children about the domestic abuse they were experiencing within their families; instead, the subject of arguments and relationships at home was alluded to each week in a less explicit way. Whatever the rights or wrongs of this decision (we could, for instance, argue that it reinforced stigma), it meant that in the evaluation, we were not able to ask the children directly whether or not the service had lessened the negative impact of domestic violence on their lives. This being the case, how could anyone claim that the service was dealing with this issue successfully?

We sought to answer this question by using a qualitative approach, including ethnography. So we sat in on one group session where we witnessed skilful and creative approaches to difficult conversations being used. Observing one session is unlikely to be enough, but limited finances meant that anything more would have been too expensive (see Baszanger and Dodier (2003) on ethnographic approaches to research). Such is the case for many evaluations. Moreover, as we now argue, outcomes are rarely single issues.

What else is going on in people's lives?

Drug and alcohol problems do not come alone; instead, they are accompanied by a range of other problems, for example, with housing, physical health, mental health, families etc. (Galvani, 2012). Because of this, the service users in the statutory evaluation were almost always accessing a range of services at the same time, as well as relying heavily on support from family and friends where this was available.

Similarly, achieving improvements in children's lives (including developing their resilience) may require sustained support over a length of time and a number of projects. Because of this, many children attended more than one group, and hence more than one project, as well as often using one-to-one support. They also accessed different services over a long period of time. The consequence was that, to a large extent, staff, referrers and parents' allegiance was to the agency, not its constituent parts (the individual services), making it impossible to either distinguish or measure one service as compared with another.

But the complexity is even greater than this. The findings from both evaluations suggest that what works best for most service users (adults and children alike) is the *range* of individuals and services that they relied on. Recent public health research suggests that evaluating 'complex interventions' is always difficult; so-called 'gold standard' randomised controlled trials (e.g. Chiapelli, 2010) may not offer any better way forward in this regard (Craig *et al.*, 2008; Mackenzie *et al.*, 2010). Research on evaluating children's services similarly shows how difficult it is to isolate specific interventions in a complex system (Cross *et al.*, 2010).

So what did we do?

Gray *et al* argue that evaluations that 'concentrate narrowly on inputs and outputs of programmes are in danger of missing vital, often illuminating information on *processes*' (emphasis in original) (2009: 284). Faced with the inevitable difficulties in measuring input, output and outcomes to any objective standard, we did what many small-scale, service evaluations in social work do: we observed practice and asked stakeholders (referrers, practitioners and service users) what they thought about the services that they used, and in doing so, attempted to say something about the processes that held them together. These then, became our 'proxies' for

measurements of effectiveness, and yet there are obvious dilemmas in this, as we will now discuss.

Discussion

We now return to the key questions posed at the outset. What is effectiveness in social work? How should it be measured? All research (qualitative and quantitative alike) is plagued with issues about reliability, validity and truth. Quantitative studies assume that the world is made up of observable, measurable facts that have an objective reality; qualitative research, in contrast, assumes no such reality, and instead focuses on issues of quality, rigour and trustworthiness (Golafshani, 2003). But how are these measurable in social work? We have shown that the 'facts' in our evaluations (certainly those in the form of numbers and 'hard' data) were questionable to say the least, found to be wholly context-specific, related as much to the 'fact' of the evaluation as to any intrinsic reality. Quality, rigour and trustworthiness may be just as illusive, and just as problematic.

What is quality and how is it to be measured?

If we are to consider quality in social work, we must first define – what do we mean by 'quality'? It may be a truism to suggest that quality is 'in the eye of the beholder', and yet this is fundamentally the case; it is inevitably a subjective experience, not an objective fact. Differing stakeholders will hold varied views on what quality is (Kettner, 2002). Moreover, divergent opinions are likely even with a single stakeholder group. Because of this, any or all of the benchmarks and standards that are used to demonstrate quality can only ever be approximations of what one individual, group or society feels to be 'good' within the limits of their own personal and professional lives.

Malley and Fernández suggest that three features make such quality measurement problematic. Firstly, services are best seen as 'performances': they have to be

experienced to be measured and verified. Secondly, performances are labour intensive, so that they can vary from day to day, and from consumer to consumer. Thirdly, performance and consumption cannot be separated from one another; it is impossible to 'disentangle the providers' influence on the quality of the service from that of users' (2010: 5). Because measurement is difficult, other structural aspects of social care tend to be measured instead, including, for example, the number of single rooms in care homes, the range of facilities etc. But these may be poor indicators of overall service user satisfaction, because it is the quality of the service user/carer relationship that is likely to be most important for users of services. Additionally, they may have little impact on user outcomes, or on the outcomes for wider family members.

This brings another complication, because although it might be relatively straightforward to demonstrate outcomes where a change occurs, outcomes are often about maintenance not change; the aim is to help someone to remain stable, or to support them as their health deteriorates until their eventual death. Malley and Fernández conclude that there is an 'attribution problem' at the heart of service quality measurement in social care: 'how to apportion the relative contributions to outcomes of services and non-service factors' (2010: 10). From this perspective, then, measuring quality is always going to be difficult, as our evaluations have demonstrated.

What about rigour and trustworthiness?

Notions of rigour and trustworthiness are at the heart of naturalistic inquiry. Lincoln and Guba (1985) argue that such an approach demands due diligence in terms of research methods, triangulation of data, and a sound knowledge of the subject-matter. In evaluation research, however, as in research literature more broadly, the concepts of rigour and trustworthiness have also often been linked to ideas of

objectivity and externality: there is a suggestion that someone who is outside of a situation will be able to bring an objective, hence 'better' view (see Stufflebeam *et al.*, 2000). It must be self-evident that neither of the evaluations discussed in this paper could be called 'external'. The Principal Investigator had worked with both agencies for a considerable time, and students had undertaken practice placements in the agencies as part of their social work training. The researchers were also 'insiders' in terms of their familiarity with the culture, jargon and every-day practice of the services. Did this then mean that the evaluations were in any way less trustworthy?

Our view is that both externality and insider-status carry advantages and disadvantages; the solution is not to give one more value than the other, but rather to unpack and problematise the impact of both. So in our evaluations, insider-knowledge, networks and contacts gave us ease of access to the social workers and referrers in the study, as well as to the service users themselves. They also, however, presented us with inevitable issues of power, because those whom we interviewed would, inevitably, have seen us as in some way implicated with the agencies themselves, potentially complicit in the agencies' own views and versions of 'truths' about the services. Brannick and Coghlan argue that 'there is no inherent reason why being native is an issue' and 'the value of insider research is worth reaffirming' (2007: 59). But taking this further, 'Whose side are we on?', asked Becker (1967). He highlighted that all research speaks from somewhere; it is vital, he argued, that we firstly admit this and then choose to speak from the view of the dispossessed. Both our evaluations took place in a context of pressure on the agencies to demonstrate their usefulness to funders and others. The bottom line was that the research team, in both instances, was aware of the potential negative repercussions that might accompany a negative evaluation - funding might be removed, services closed, and service users' lives negatively impacted. We were not, then, neutral bystanders, but instead we *cared* about the evaluation and its

consequences. And rather than pretending neutrality, we preferred to bring a reflexive understanding to this reality (Finlay and Gough, 2003) by discussing this openly with our interviewees, and by presenting this ever-present dilemma in our final project reports.

Conclusion

We began by asserting the usefulness of evaluation for individual practitioners and for agencies. However, we have argued that it is also a neoliberal exercise that gives agencies the machinery with which to compete against one another within a quasi-market of care; it is, in effect, a managerialist tool that provides value to some activities over others, most especially those that can be measured easily, or that reflect the values and voices of the mainstream (Gray *et al*, 2009). At the same time, alternative ways of knowing and being are not given prominence, and voices of dissent are not heard (Webb, 2001).

Our view, in conclusion, is that it does not need to be this way. On the contrary, by affirming the importance of evaluation, there is a possibility of re-engaging with the agenda and offering an alternative vision. In doing so, we would like to propose a critical paradigm, which understands that social work practice is uncertain, messy and complex (Fook, 2007) and which sees the role of evaluator as a 'critical friend' (Balthasar, 2011) working alongside agencies and service users to co-create an evaluation that is 'owned' by everyone (Whitmore, 2001). Such an approach will be explicit about the political context in which the evaluation is located (Frost, 2002), while giving attention to different stakeholders' knowledges (Pawson and Tilley, 1997) and valuing practice wisdom and service user experience (Gray *et al.*, 2009).

A critical paradigm might look like this:

1. Knowledge about the evaluation will be co-created from the outset, with agency staff and service users working with academic researchers to design what the evaluation should be about, and how it should be carried out.
2. The researcher will see her/himself as working with, not against the agency, as a 'critical friend'.
3. The evaluation will typically use a mix of methods, reflecting the idea of valuing different and multiple realities, perspectives and experiences.
4. Methods will seek to address issues of power, for example, by supporting service users and/or practitioners to conduct all or parts of the evaluation, and the researcher becoming mentor or peer educator.
5. Analysis will be developed along the way, and feedback discussion will likewise take place at key points, not just at the end of the evaluation.

By using such an approach, we can reclaim evaluation as a practice that is worthwhile for social work practitioners, agencies and those who use services, rather than something that is imposed from outside as a NPM construction. It becomes part-and-parcel of a self-reflexive, critical practice as described almost 20 years ago by Shaw and Shaw (1997), when they urged us to keep social work 'honest'.

References

- Ayre, P. (2001) 'Child protection and the media: lessons from the last three decades', *British Journal of Social Work* 31: 887 - 901.
- Balthasar, A. (2011) 'Critical friend approach: policy evaluation between methodological soundness, practical relevance and transparency of the evaluation process', *German Policy Studies*, 7(3): 187-231.
- Bamberger, M., Rugh, J. and Mabry, L. (2012) *Real World Evaluation*, 2nd edition, Sage, Thousand Oaks, CA.

- Banks, S. (2011) 'Ethics in an age of austerity: Social work and the evolving New Public Management'. *Journal of Social Intervention: Theory and Practice*. 20(2) pp.5–23. DOI: <http://doi.org/10.18352/jsi.260>
- Barnes, M. (1994) 'Objective science or social interaction? Researching users' views of services', *Research, Policy and Planning*, 12(2): 1-3.
- Baszanger, I and Dodier, N. (2003) *Ethnography: relating the part to the whole*, in Silverman, D. (ed.) *Qualitative Research. Theory, Method and Practice*, London: SAGE, pp9-34.
- Becker, H. (1967) 'Whose side are we on?' *Social Problems* 14(3): 239-247.
- Best, J. (2001) *Damned Lies and Statistics. Untangling Numbers from the Media, Politicians and Activists*. University of California Press, Berkeley.
- Boote, D.N. and Beile, P. (2005) 'Scholars before researchers: On the centrality of the dissertation literature review in research preparation', *Educational Researcher* 34(6) pp3-15.
- Cheetham, J., Fuller, R., Mclvor, G. and Petch, A. (1992) *Evaluating Social Work Effectiveness*, Buckingham: Open University Press.
- Cheetham, J., Mullen, E., Soydan, H. and Tengvald, K. (1998) 'Evaluation as a tool in the development of social work discourse', *Evaluation*, 4(1): 9-24.
- Cheetham, J. and Mansoor, A.F.K. (1998) *The Working of Social Work*, London: Jessica Kingsley.
- Craig, P., Dieppe, P., Macintyre, S., Mitchie, S., Nazareth, I., and Petticrew, M. (2008) 'Developing and evaluating complex interventions: The new Medical Research Council guidance'. *British Medical Journal*, 337(7676) pp979-983.
- ANON (2009)
- ANON (2014)
- ANON (2016)

- Cross, S. Hubbard, A. and Munro, E. (2010) *Reclaiming Social Work*, London Borough of Hackney Children and Young People's Services, Human Reliability and LSE, London.
- Everitt, A. and Hardiker, P. (1996) *Evaluating for Good Practice*, Birmingham: BASW.
- Ferguson, I. and Woodward, R. (2009) *Radical Social Work. Making a Difference*, Bristol: Policy Press.
- Finlay, L. and Gough, B. (eds) (2003) *Reflexivity. A Practical Guide for Researchers in Health and Social Sciences*, Oxford: Blackwell.
- Fischer, J. (1978) *Toward Effective Casework Practice: An Eclectic Approach*, McGraw-Hill, New York.
- Fisher, M. (1999) 'Social work research, social work knowledge and the research assessment exercise', in Broad, B. (ed.) *The Politics of Social Work Research and Evaluation*, Birmingham: Venture Press.
- Fook, J. (2007) 'Uncertainty: The Defining Characteristic of Social Work?', in Lymbery, M. and Postle, K. (eds) *Social Work. A Companion to Learning*, London: Sage.
- Frost, N. (2002) 'Evaluating practice', in Adams, R., Dominelli, L. and Payne, M. (eds) *Critical Practice in Social Work*, Palgrave, Basingstoke, pp46-54.
- Galvani, S. (2012) *Supporting People with Alcohol and Drug Problems. Making a Difference*. Bristol: The Policy Press.
- Golafshani, N. (2003) 'Understanding reliability and validity in qualitative research', *The Qualitative Report*, 8 (4): 597-607.
- Gomm, R., Needham, G. and Bullman, A. (2000) *Evaluating research in health and social care*, London: Open University and Sage.
- Gray, M., Plath, D. and Webb, S.A. (2009) *Evidence Based Social Work. A Critical Stance*, London: Routledge.
- Gruening, G. (2001). 'Origin and theoretical basis of New Public Management'. *International Public Management Journal* 4: 1-25.

- Hart, C. (1998) *Doing a Literature Review: Releasing the Social Science Research Imagination*, London: SAGE.
- Harvey, D. (2005) *A Brief History of Neoliberalism*, Oxford: Oxford University Press.
- Hayes, D. and Devaney, J. (2004) 'Accessing Social Work Case Files for Research Purposes. Some Issues and Problems', *Qualitative Social Work* 3(3) pp313-333.
- ANON (2014)
- Johnson, R.B., Onwuegbuzie, A.J. and Turner, L.A. (2007) 'Toward a Definition of Mixed Methods Research', *Journal of Mixed Methods Research* 1(2): 112-133.
- Jones, C. (2005) 'The neo-liberal assault: voices from the front line of British state social work, in Ferguson, I., Lavalette, M. and Whitmore, E. (eds) *Globalisation, Global Justice and Social Work*, London: Routledge.
- Kettner, P.M. (2002) *Achieving Excellence in the Management of Human Service Organizations*, Boston: Allyn & Bacon.
- Lincoln, Y.S. and Guba, E.G. (1985) *Naturalistic Inquiry*, Newbury Park, CA: Sage.
- Lishman, J. (ed.) (1984) *Evaluation (Research Highlights in Social Work)*, London: Jessica Kingsley.
- Mackenzie, M., O'Donnell, C., Halliday, E., Sridharan, S. and Platt, S. (2010) 'Evaluating complex interventions: One size does not fit all', *British Medical Journal* 340(7743) pp401-403.
- Malley, J. and Fernández, J.-L. (2010). 'Measuring quality in social care services: theory and practice'. *Annals of public and cooperative economics*, 81(4). pp. 559-582. ISSN 1370-4788 DOI: 10.1111/j.1467-8292.2010.00422.x
- Marsh, P. and Fisher, M. (2005) *SCIE Report 10: Developing the evidence base for social work and social care practice*, London: SCIE
- <http://www.scie.org.uk/publications/reports/report10.asp>
- Mayer, J.E. and Timms, N. (1970) *The Client Speaks: Working Class Impressions of Casework*, London: Routledge and Kegan Paul.

- MacDonald, G. and Sheldon, B. (1992) 'Contemporary studies of the effectiveness of social work', *British Journal of Social Work*, 22(6): 614-43.
- Oliver, P. (2006) 'Purposive sampling', in Judd, V.(ed.) *The SAGE Dictionary of Social Research Methods*, London: Sage.
- Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*, Sage, London.
- Rees, S. and Wallace, A. (1982) *Verdicts on Social Work*, Edward Arnold, London.
- Richmond, Mary Ellen (1913) *A study of nine hundred and eighty-five widows known to certain charity organization societies in 1910*, New York: Charity Organization, Russell Sage Foundation.
- Richmond, M.E. (1917) *Social Diagnosis*, New York: Russell Sage Foundation.
- Rogowski, S. (2010) *Social Work. The Rise and Fall of a Profession?* Bristol: Policy Press.
- Scottish Government (2008) *The road to recovery: a new approach to tackling Scotland's drug problem*. Edinburgh: Scottish Government.
- Scottish Government (2009a) *Changing Scotland's relationship with alcohol: a framework for action*. Edinburgh: Scottish Government.
- Scottish Government (2009b) *An Outcomes Toolkit for Alcohol and Drugs Partnerships (Version 1) provides guidance on working towards an outcomes approach for 'Alcohol and Drug Partnerships' across Scotland*, Edinburgh: Scottish Government.
- Shaw, I. (2011) *Evaluating in Practice*, Ashgate, Farnham.
- Shaw, I. and Norton, M. (2007) *Kinds and Quality of Research in Social Work*, London: Social Care Institute for Excellence.
- Shaw, I. and Shaw, A. (1997) 'Keeping social work honest. Evaluating as profession and practice', *British Journal of Social Work*, 27(6): 847-69.
- Stufflebeam, D.L., Madaus, G.F. and Kellaghan, T. (eds) (2000) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation (Evaluation in Education and Human Services)* 2nd edition, New York City, Springer, NY.

- Trinder, L. (1996) 'Social work research: the state of the art (or science)', *Child & Family Social Work* 1(4): 233-42.
- Trinder, L. (2000) 'Evidence-based practice in social work and probation', in L. Trinder, L. and S. Reynolds (eds) *Evidence-based Practice: A Critical Appraisal*, Oxford: Blackwell Science.
- Webb, S.A. (2001) 'Some considerations on the validity of evidence-based practice in social work', *British Journal of Social Work* 31: 57-79.
- Whitmore, E. (2001) "' People listened to what we had to say": reflections on an emancipatory qualitative evaluation', in I. Shaw and N. Gould (eds) *Qualitative Research in Social Work*, London: Sage.
- White, S., Wastell, D., Broadhurst, K. and Hall, C. (2010) 'When policy o'erleaps itself: the tragic tale of the Integrated Children's System', *Critical Social Policy* 30: 405-429.
- Wilcox, A., with Hoyle, C. and Young, R. (2005) 'Are randomised controlled trails really the 'gold standard' in restorative justice research?' *British Journal of Community Justice* 3(2): 39-49.
- Wilkinson, S. (2004) Focus group research, in Silverman, D. (ed.) *Qualitative Research. Theory, Method and Practice*, Sage, London, pp177-199.