Book: Responsibility and governance: the twin pillars of sustainability
Chapter 6: Wond, T. Evaluation for what purpose? Findings from two stakeholder groups
Editors: David Crowther, Shahla Seifi & Tracey Wond
Springer Singapore.

# Evaluation for what purpose? Findings from two stakeholder groups

**Tracey Wond, University of Derby, UK**

## Abstract

A host of reasons exist for the pursuit of evidence in the public sector, including to support good governance and policy development. As the expectations for program evaluation from policymakers have evolved, so too has evaluation practice and a great deal of experimentalism has ensued. There is a risk that these developments and the complexity inherent in them, may lead to conflicting expectations about why program evaluation is done, or even a loss of purpose. This prompts the meso-level analysis of two types of stakeholders in a governance network, explored in this chapter. This chapter presents the findings of an ongoing study which explores the perceptions of program evaluators and policy implementers towards the purpose of evidence.

The findings suggest that program evaluators and policy implementers have divergent expectations of why and how evaluation data might be used. The findings suggest that program evaluators aspire to support change and enhance the policy domains they serve, whereas policy implementers perceive program evaluation as serving a more governance-/management-orientated role.

The chapter demonstrates the complexity of both program evaluation and policy and may have implications for the twin pillars of governance and responsibility at the heart of the book. If governance and responsibility are the twin pillars of sustainability then the complex networks of relationships, expectations, values, and outcomes may need to be considered. The findings also have implications for evaluation commissioners and practitioners, demonstrating the need for the purpose and expectations of program evaluation to be agreed early. The use of program evaluation as a symbolic, aesthetic or structural mechanism also emerges, prompting opportunity for further research, for instance, to explore legitimacy and program evaluation.

*Keywords: evaluation utilization; program evaluation; governance; evidence-based policy and practice; complexity*

## Introduction

Historically, the development of evaluation has been linked to public reform and the emergence of the New Public Management (NPM) ideology (Shaw, 1999; Taylor, 2005). NPM led to public sectors across the world engaging heavily with evidential mechanisms such as evaluation, since evaluation provided a means to ensure good governance, value for money, efficiency and accountability (Boaz and Nutley, 2003; Bovaird and Loeffler, 2007; Stern, 2008). Whilst the 'the epistemic bubble surrounding NPM' has now burst (Brinkerhoff and Brinkerhoff, 2015, p.223), the trend for evidence and evaluation data has not.

However, tension towards the role and function of evaluation has mounted over several decades, and challenges exist in how evaluators perform evidential and governance roles. These challenges could be broadly defined as *definitional*, *functional* and *contextual*. Evaluation faces *definitional* challenges such as the extensiveness of the evaluation concept which has proven problematic and confusing (Weiss, 1972; McKie, 2003). Evaluation is not just a research activity but also a profession. As a research activity, evaluation extends from basic monitoring data, to auditing (a regulated profession in its own right), to more diverse and complex methodological approaches and design. Evaluators may undertake evaluation activity full-time, in internal or external capacities, whilst others often undertake evaluation as part of a wider role (academics undertaking evaluation for instance). At a *functional* level, evaluation, including program evaluation, is vast, evolving to demonstrate responsible public spending, success, and identify what works - often with sophisticated methodologies (Bristow et al, 2015). The extensive and vast scope of evaluation complicates expectations that others may have of it. Adelman (1996, p.295) acknowledges the tense and contradictory relationship between evaluation and policy-makers: 'evaluators want to influence policy-making, but few were willing to participate in the process of decision-making; that was the responsibility of policy-makers'. Evidence discourse continues to develop and this, in part, represents a *contextual* challenge for evaluation. For instance, 'growing disagreement and confusion about what constitutes sound evidence for decision making' has been noted, making the role of an evaluator difficult (Donaldson et al, 2009, p.12). Further, there is increasing acknowledgement that public policy, governance arrangements and evaluation co-exist in complex environments (Walton, 2016). It is reasonable to assume that, as models of public administration become more polycentric and a greater diversity of stakeholders become involved in policy development, decision-making, and governance, that this complexity will intensify (Evers and Ewert, 2012). This chapter acknowledges these complex systems.



This chapter, and the research underpinning it, responds to recognition that a dearth of studies have explored the practice of evaluation in general, and program evaluation in particular (Fitzpatrick et al, 2009). Pawson and Tilley (1997, p.24) suggested, over twenty years ago, that a synthesis of evaluation theory was necessary, given that evaluation had been 'tossed back and forth on a sea of favourable and ill tides'. But still, little exploration has come (with the exception of many methodological contributions). Seppanen-Jarbela (2003, p.76) asserts that, 'there is an obvious need to rethink why, what for, and who for evaluation data is collected', and this chapter explores the 'what for' part of this assertion. The study, on which this chapter is based, sought to identify and compare the perceptions of program evaluators and policy implementers towards the purpose and use of evidence. It is based on semi-structured interviews with 19 program evaluators and 10 policy implementers (managers and senior personnel of publicly funded programs) in the UK. The study therefore, explores evidence use at the meso-level in a governance network context.

The twin pillars construct, central to this book, supposes a relationship between governance, responsibility, and sustainability. Evaluation and the role of evidence have been linked heavily to the concepts of responsibility and governance represented in these twin pillars, particularly within public administration and evaluation sciences literature (Bovaird and Loeffler, 2007; Davies, 1999; Hansson, 2006; Newcomer, 1997; Stern, 2008; Taylor, 2005). The findings presented herein encourage those exploring the twin pillars construct to consider the changing modes of governance and to acknowledge the increasingly complex systems in which program evaluation and policy concerns, such as environmental sustainability, reside.

Next, a literature review introduces evaluation and critically explores the challenges surrounding it. A methodology then outlines how the study was designed and implemented. The findings and discussion section follows and explores the findings and attempts to make sense of them. Finally, the chapter is concluded with consideration for the implications and future directions of the study.

## Literature Review

### *Introducing Evaluation*



Evaluation is the 'systematic study of the behaviour of groups of individuals in various kinds of social settings, using a variety of methods' (Christie and Fleischer, 2009, p.21). Whilst stemming from the social sciences, evaluation extends several disciplines, attracting the label of a 'metadiscipline' (Picciato, 1999. p.7) and 'transdiscipline' (Scriven, 1996, p.402). Indeed, evaluation has become commonplace in public settings such as health, education and international development. Upon exploring the various typologies, definitions, and interpretations of evaluation its extensiveness can be observed (Berk and Rossi, 1990; Husbands, 2007; Shaw and Faulkner, 2006). Stecher and Davis (1988, p.23) summarise this position, 'there is no single, agreed upon definition of evaluation…there are a number of different conceptions about what evaluation means and how it should be done'. This chapter focuses on program evaluation, defined as, 'a social science activity directed at collecting, analysing, interpreting, and communicating information about the workings and effectiveness of social programs' (Rossi et al, 2003, p.3). Programs often attempt to introduce interventions or be focused on prevention or promotion of particular activity (Chen, 2014).

Chelimsky and Shailesh (1997) identify that evaluation can serve purposes of 'accountability', 'knowledge', and 'development'. Others have broadly supported these categories, identifying the need for evaluation to support policymakers to learn (Husbands, 2007; Shaw and Faulkner, 2006) and to assure governance (Shaw and Faulkner, 2006). Although, in practice, the extent to which evaluation is used for these purposes differs. These varying purposes present scope for confusion and conflicting expectations (Weiss, 1972; McKie, 2003).

The potential for evaluation to serve a learning or developmental purpose are relatively well acknowledged (Fitzpatrick et al, 2009; McCoy and Hargie, 2001; McKie, 2003; Shaw and Faulkner, 2006; Weiss, 1972). This includes developing program theory, influencing the design of policy interventions, providing ongoing feedback, and influencing future interventions. Indeed, several evaluation constructs have emerged to distinguish this learning purpose. Within

Scriven's (1980, p.6) formative/summative dualism, formative evaluation identifies the provision of feedback 'to improve something', and places the provision of knowledge to policy decision-makers as important. However, many have contended this learning purpose (Bovaird and Davis, 1996; Iriti et al, 2005), in particular questioning the impartiality, distance and independence of evaluation. The expansive nature of evaluation (extending to audit, performance measurement, process evaluation etc.) compounds such contention:

> When we had no interest in changing anything we had less need to explain – it would suffice to assess performance…evaluation has moved upstream to become involved in policy analysis and programme design and downstream towards implementation and change management
>
> (Stern, 2008, p.251)

The movement of evaluation (supposedly) 'downstream', towards development and implementation is not necessarily a new role for evaluation and resonates with its historical roots in supporting US reform.

### *Evaluation for governance*

Governance has become a catch-all term (Frederickson, 2005) and it is important to define what it is, in order to understand what is implied by the notion of 'evaluation for governance' described here. Klijn (2008, p.507) identifies four main definitions of governance, 'good/corporate governance' (relating to the fair and proper operation of government); 'new public management' (embedding performance measurement and accountability mechanisms for those delivering public services); 'multi-level' and/or 'inter-governmental' (embracing the multiple layers and hierarchies of public organisations); and, 'network governance' (the need to manage across networks and complex systems, that often cross boundaries are involve multiple actors). By Klijn's (2008) definition, governance is pervasive, difficult to bound and scalable - typical of a complex system (CECAN, 2018). Similarly, Crowther et al (2017) identify good governance as comprising, 'transparency', 'accountability', 'responsibility' and 'fairness'. This appears to integrate Klijn's (2008) notion of new public management, which was kept separate in Klijn's definitions. The role of evaluation in supporting each of these four principles is well supported, as table 1 demonstrates.

<Table 1 here >

Public management reform has driven much of the evaluative activity that we observe today (Hansson, 2006; Head, 2008; Henkel, 1991; Taylor, 2005). Efficiency, control, value for money and accountability were central components of 'New Public Management' (NPM) (Hansson, 2006), and evidence to demonstrate the achievement of such outcomes was necessary (Davies, 1999; Bovaird & Loeffler, 2007; Stern 2008). Evaluation became a 'key entry in the lexicon of new public management (Taylor, 2005, p.602) and a measurement-driven culture ensued in public services in many parts of the world (Kettl, 2005; Klijn, 2008; Taylor, 2005). Attempts by public services to be more responsible, strategic and efficient have reinforced the evidence-based policy and practice concept, and the need for evaluation for evidence - 'to promote accountability and control' (Sanderson, 2000; Shaw, 1999). Further, evaluation has played a key role in embedding and legitimising neo-liberalism (Giannone, 2016). Greater public voice and media scrutiny have also pressured public services to demonstrate responsible stewardship of 'taxpayers' money' (Barbier, 1999, p.378).

The changing face of the public sector and form that governance takes may also have implications for evaluation. Greater private-public partnerships have stimulated a fundamental rethink on how governance is assured and the notions of 'co-governance' and 'new public governance' have emerged as a result (Osborne, 2000; Theisens et al, 2016). 'New forms of horizontal governance' (Klijn, 2008, p.506) have embedded more polycentric and participative forms of public sector decision-making and are structured around greater citizen involvement. Evaluation has responded with more participative methodological approaches (Plottu and Plottu, 2009). As more stakeholders become involved in governance and evaluation, there is a risk that it becomes increasingly difficult to satisfy the expectations and stakeholder claims of the many. Stakeholder identification and salience theory suggests that not balancing the claims of various stakeholders (after assessing various attributes to determine salience) may result in harm to the organisation (Neville et al, 2011).

### *Challenges Facing Evaluation*

A growing number of evaluation theorists have noted the challenges that face evaluators. These challenges emerge at various stages of the evaluation process and commonly relate to the use of evaluation, the effectiveness of certain methodologies and funding regimes (which in particular can suffer short-termism), and tensions during evaluation interactions.

The under-utilisation and effectiveness of evaluation findings has frustrated both evaluation communities and those who fund them, expounding criticism that evidence-based policy and practice (as a mechanism that evaluation feeds) is ideological, flawed and failing (Parkhurst, 2017). Despite significant public expenditure being committed to construct such evidence bases (see for instance National Audit Office, 2013), the use of evidence to inform policy intervention is sporadic, and there are increasing accounts of the underuse and misuse of such evidence (Weiss, 1993; Wond, 2017). EBPP has endured a great deal of criticism, and despite maturing as a concept, in practice, it has struggled to become fully institutionalised or legitimated in many areas of public policy. The failure of negative evaluation reports, including UK Brexit impact assessments, to be disclosed in a timely manner are examples of this. Chelimsky (2008) noted that evaluator contracts have been known to be terminated when negative findings have been reported. Indeed, Hanberger (2006) notes that evaluation has the potential to be undemocratic – used to cover program failures and further self-interests. The timeliness of evaluation findings has also been posed as problematic for two reasons, (1) the time that program impact takes to fruit; (2) policy interventions have often evolved before summative evaluation findings are published (Weiss; 1972; Wond and Macaulay, 2010).

A host of elaborate methodologies and perspectives of what makes good evidence have emerged (reigniting the quantitative/qualitative paradigm war in doing so), with the potential of further alienating those policymakers, for whom the evidence is meant to support (Bristow et al, 2015). Walton (2016) suggests that as policy systems become more complex, buy-in across governance networks, rather than further technical sophistication is necessary.

Many of the interactional complexities inherent in evaluation have been integrated into the concept of excessive evaluation anxiety, which Donaldson et al (2002) explored in the program evaluation context. According to Donaldson et al (2002, p.261), 'the experience of being evaluated, critiqued, or judged commonly results in an emotional reaction of uneasiness, uncertainty, or apprehension'. XEA is a negative anxiety resulting from the evaluation process and experienced by evaluation participants including program staff. There are many ways in which XEA appears to have manifested itself in previous studies, including access issues, inaccurate data reporting, conflict, and reluctance to participate or cooperate (see also Bechar and Mero-Jaffe, 2014). The manifestation of evaluation anxiety in program settings is considered to be deeply unproductive, and can render 'evaluation data and findings worthless', and 'undermine the credibility of evaluators' (Donaldson et al, 2002, p.263).

***Complex Evaluation Systems***

Evaluation occurs within inherently political settings (Chelimsky, 2008; Mohan and Sullivan, 2006; Pawson and Roe, 1982; Taylor, 2005; Tilley, 1997; Weiss, 1993). The publicly funded nature (for instance, through public sector services, government funding, and donor-giving in charitable settings) of many evaluations complicates the evaluation setting.

According to Stern, 'evaluators now find themselves necessarily connecting to a host of 'policy communities', 'civil society actors', 'excluded groups', 'stakeholders', 'commissioners', 'professionals' and 'academic disciplines' (2008; p.249). The diversity of stakeholders and the varying expectations and interests that stakeholders hold further complicates the evaluation setting (Mohan and Sullivan, 2006; Stufflebeam, 2001). In a program evaluation setting, stakeholders might include program managers, program delivery staff (these two groups form our policy implementers in this study), beneficiaries, policy-makers, commissioners, the evaluator, and academics (McCoy and Hargie, 2001; O'Brien et al, 2010; Stern, 2008; Wond, 2017). Whilst this makes for a complicated environment, Mohan and Sullivan (2006) highlight that stakeholder diversity can be beneficial, particularly to support evaluators to appreciate the varying perceptions that exist (beyond their own).



Program evaluators have the potential to influence decision-makers through their findings, for instance by stimulating program improvements, or providing evidence that supports or hinders continuation grant applications (Feinstein, 2002; Pollitt, 2003). This potential influence could lead to tension and influence trust within the program evaluation relationship (Roe, 1982; Wond, 2017): 'sometimes stakeholders are hypersensitive to evaluators because they perceive they have much to gain or lose based on the success of failure of the program' (Donaldson et al, 2002, p.267). Vice versa, evaluation stakeholders can also significantly influence the evaluation and evaluator (Patton, 1997). Davies (1999; 155) warns that data may be 'hijacked by senior management or political masters'. References to power in the evaluation relationship are relatively limited and are at times subtle, but observable (Diamond, 2005; Tourmen, 2009), affecting methodological choices, evaluation utilisation and the health of evaluation relationships. Pawson and Tilley note there has been a 'failure to appreciate the asymmetries of power in the evaluation relationship' (1997, p.20). Donaldson et al (2002, p.267) suggest that evaluators should be mindful of this 'program psychologic' when presenting evaluation findings.

Complexity theory acknowledges the complex systems in which activity such as evaluation occurs, making it worthy of some introduction. Complexity is 'a form of order that emerges when certain sets of things interact in certain ways with one another' (Castellani and Hafferty, 2009: 123). Complexity theory began in the physical sciences (Walton, 2016) and gradually expanded into fields including management, organisation, and public administration sciences. Complexity has been described as very many things including a methodology, philosophy and theory (Haynes, 2008; Walton, 2016). There are many characteristics inherent in complex systems including the cross-boundary nature of activity and issues (Meek, 2014), non-linearity (Gilbert, 2017), boundlessness (CECAN, 2018), uncertainty and unpredictability (Meek, 2014; CECAN, 2018). According to Gilbert (2017, p.5): 'the characteristics of a complex system, using the term in its technical sense, are that it consists of many units that interact and that as a result, the behaviour of the system as a whole is more than just the aggregation of the behaviours of the units'. There have been several attempts to explore and apply complexity theory to (program) evaluation settings - as a methodology to undertake evaluation, as well as an approach to meta-evaluate (perform an evaluation of evaluation)(Gilbert, 2017; Haynes, 2008; Morrell, 2010; Walton, 2016). There is increasing awareness that the complex policy systems in which evaluation operates requires such an approach (CECAN, 2018; Walton, 2016). As co-governance arrangements proliferate public organisations in the UK, New Zealand and farther afield, further consideration for the increasing complexities and approaches to manage these may be necessary (Duncan and Chapman, 2012; Walton, 2016).

## Methodology

The study explored in this chapter provides a meso-level exploration of two types of communities (program evaluators and policy implementers) participating in public governance networks. Klijn (2008, p.511) defines governance networks as 'public policy making and implementation through a web of relationships between government, business and civil society actors'. To clarify the association between public policy, programs and program evaluation, public policy intervention is often piloted and delivered through the form of programs. These programs are often evaluated (program evaluation) in order to explore how they worked and whether they were effective. Both policy implementers and program evaluators were considered to be actors participating in decision-making in these governance networks. Prior studies have supported the notion of evaluation residing within governance networks (Walton, 2016).



The study involved semi-structured interviews with 19 practicing program evaluators (9 female and 10 male) who undertake evaluation in various capacities (academics, evaluation consultants, internal program evaluators) in public and charitable settings. Interviews with 10 policy implementers (managers/staff in or supporting public programs) were also conducted. Both stakeholder groups represented a range of policy areas (e.g. health, education, foreign aid, enterprise support). The interviews were undertaken as part of a wider ethnographic study and through contextualised questioning looked to understand respondents wider experiences of program evaluation, challenges they felt limited evidence use, and how they used or perceived the use of evaluation evidence.

An informal interviewing approach, with planned questions but without a strict interview guide (Brinkmann, 2013), was adopted, and questions were generally broad and open-ended. A phenomenological research and interview approach were adopted, embedded key principles such as 'natural attitude' (Bevan, 2014). Phenomenological research approaches are known for their appreciation of complexity and the interconnectedness of aspects influencing experience (Mason, 2002). An approach to 'bracketing' (Bevan, 2014; Husserl, 1970), the conscious restraint of the researcher in respect of their prior personal knowledge, was a more difficult design decision. The researcher attempted to bracket their existing knowledge using Tufford and Newman's (2010) conceptual framework for bracketing, which draws attention to numerous stages within a research study that bracketing could be considered. By being aware and through reflection in the researcher's field notes journal (a continuation of the journal used for the ethnographic element of the wider study), attempts were made to bracket and mitigate personal knowledge and presuppositions during study conceptualisation and design, data collection, data analysis and writing. Nonetheless, the researcher's prior experience did support meaning to be assigned to some areas of the discussion, and to iteratively build upon answers from respondents where further clarification or opportunities for greater depth appeared – without a knowledge of the program evaluation context this might have been more difficult to achieve. The researcher's prior knowledge of performing program evaluation supported the interpretative validity of the research, yet contradicted the approach to bracketing associated with phenomenological interviewing. Interviews were predominantly administered face-to-face, but some online and telephone interviews were also conducted.

The researcher's own involvement with evaluation societies and networks in the UK and Europe supported access to program evaluators, using an opportunity sampling approach. Further networking and involvement in project settings led the researcher to access policy implementers for this research. In terms of external validity, the study population comprised of UK and European program evaluators, and UK policy implementers, thus providing a rather UK-centric view of program evaluation – a limitation that should be noted. Interviews were transcribed contemporaneously for the most part. Data were coded and analysed using NVIVO, respondent attributes were not assigned or analysed given the opportunistic sampling approach, size and diversity of the sample.

The findings presented below are abridged, since the wider study focused on many more aspects of evaluation perceptions and practice. A subset of the data, which related to the purpose and perceived use of program evaluation, is analysed here.



## Findings and Discussion

### *Making the world better: the moral mission*

Many program evaluators emphasised that they hoped their work would be used to (in the words of one respondent), '*make a difference'*, and (in the words of another), '*make the world a better place*'. There appeared to be a strong association with program evaluation for improvement and learning. Evaluators closely associated their work with enhancing outcomes for beneficiaries: *'to improve the lives of beneficiaries'*, '*to provide learning on what worked to make future activity better'*, and '*to make a difference'*. For instance, one evaluator in a health setting described how she wanted to find that the intervention worked, to reduce risks and prevent harm to patients – if the intervention didn't work she wanted to aid learning to improve future attempts. Therefore, a strong moral purpose was evident amongst program evaluators. This might demonstrate an implicit perceived or felt connection between the program evaluator and planned policy intervention. The sense of program evaluation for improvement resonates with evaluation discourse, where some have identified this improvement-focus (McKie, 2003; Shaw and Faulkner, 2006; Fitzpatrick et al, 2009).

In contrast, only two of the ten policy implementers referenced the potential for program evaluation to be used for improvement, with terms such as '*feed(ing) the evidence-base*', and *'showing us what works'* being used. As the findings above demonstrate, there was some disparity between how program evaluators hoped their work would be used, and how policy implementers perceived the use of program evaluation. The motivations of program evaluators to fulfil a 'moral mission' emerged as an interesting finding and may be worthy of further exploration. Since there was a lack of recognition of this by policy implementers, a starting point for realising this motivation may be greater communication of this aspiration to effect change by program evaluators.

Despite this ideological view of evaluation depicted by program evaluators, there was recognition by both program evaluators and program implementers that program evaluation use was often limited. This strongly aligns with the evaluation and evidence-based policy and practice discourse.

***Evaluation for Governance***

The majority of policy implementers perceived that program evaluation should play a heavy governance and monitoring role, and they spoke of evaluation as instrumental in proving targets and assuring responsible spending. Policy implementers spoke of program evaluation functioning *'to capture how many beneficiaries there were and if we hit our targets'*; '*to report that the money was spent properly'*; '*so we can monitor what we do'*, and *'(evaluation)…allows us to draw down the next lot of funding'*. Program evaluators also recognised this governance role and the pursuit of '*the usual monitoring 'stuff'*' emerged in 11 of the 19 interviews (and more so amongst internal evaluators).

As policy implementers discussed program evaluation, they referred to *tracking* performance, to '*monitoring'* and to evaluate for '*proof'*. A more quantitative connotation of program evaluation also appeared, with the exception of mention by three policy implementers that evaluators helped programs to generate case studies and present good news. This finding suggests that policy implementers associated program evaluation with cursory measures of performance.

Evaluation for accountability and transparency emerged through many of the interviews and a clear governance role for program evaluation was recognised by policy implementers. This supports evaluation discourse which suggests that evaluation has a role in assuring governance. Albeit, many of the responses in this study referred to the basic monitoring function underpinning program evaluation, as opposed to more elaborate or technical modes of program evaluation. The evaluation community may be disappointed to note such findings, since such perceptions (by policy implementers) appear to over-simplify the knowledge and skills needed to undertake evaluation. Since the study was conducted across a range of sectors and with no fixed definition of 'evaluation' or 'program evaluation' set to aid responses, this finding should be taken cautiously and requires further exploration. The confirmation of an evaluation for governance role also supports the decision to consider the two communities under study here (program evaluators and policy implementers), as participating in 'governance networks'.

It was interesting to note that, equity or fairness, despite being a component of governance (Crowther et al, 2017), did not feature in any of the 29 interviews. Program evaluation could, conceivably, function to ensure fair and equitable allocation of program resources to beneficiaries, and so the absence of this role was thought-provoking.

### *Means to an end?*

There was recognition, by both program evaluators and policy implementers, that evaluation reports were under-utilised. Program evaluators spoke of evaluation reports '*gathering dust'*, '*abandoned*', *'lost'* in office drawers, and '*unread in an inbox somewhere'*. Policy implementers spoke less of evaluation reports, as a product of program evaluation. Instead, the importance of monitoring data repeatedly emerged, as did evaluation as a mechanism to break through key stage-gates (to borrow from project management terminology). For instance, to ensure the continued release of funding (*'to draw down the next lot of funding'; 'to secure year three'*) and demonstrate targets had been met within particular reporting periods.

Program evaluators, in emphasising evaluation reports, appeared to associate evaluation as a product or deliverable (Scriven, 1980); they also spoke more about the future impact of their work (the moral mission noted earlier). Policy implementers could be seen to describe evaluation as more of a process, mechanism or instrument to support other ends (e.g. such as securing more funding, legitimation etc.).

### *Symbolic versus structural use of evidence*

Evaluation was frequently spoken of as supporting programs/policy interventions to be '*seen to'* deliver or achieve certain outcomes or act in particular ways. Similarly, several policy implementers spoke of evaluation as a *'tick-box'* exercise that needed to be done (one program evaluator also recognised that evaluation was perceived in this way). As such, a symbolic and aesthetic role for program evaluation was recognisable. There is resonance here to legitimacy theory, and in particular strategic legitimacy that recognises organisations exaggerating or even falsifying claims of compliance in order to be seen to act in accordance with societal norms (Suchman, 1995). Much attention has been paid to strategic legitimacy in respect of social or environmental reporting for instance.



## Conclusion

The findings offer an empirical account of the challenges facing stakeholders concerned with program evaluation - a relatively neglected area of governance, public administration, and evaluation literature. There are several implications from the findings presented, particularly given the contradictory perceptions of the use of program evaluation. These study findings have implications for policy, practitioner, and scholarly communities.

One theme that emerged strongly was the divergence in perceptions between policy implementers and program evaluators, who appeared to differ on matters such as the use of evaluation reports, and the purpose of program evaluation as a whole. Such incongruence between the supply of program evaluation (by program evaluators) and demand for program evaluation (by policy implementers) echoes concerns in the literature that evaluation and policy are evolving away from one another (Donaldson et al, 2009). Misaligned action and intention between program evaluators and policy implementers may affect the position, legitimacy and overall effectiveness of the evaluation function. The implications of such incongruence are outlined in both stakeholder and complexity literature in other contexts. Stakeholder theories acknowledge that cooperation and collective action supports the salience of particular stakeholders (Ali, 2017). From a complexity lens, a 'divergence in the values and assumptions' of stakeholders is typical of a complex policy system (Walton, 2016, p.76; Meek, 2014).

Despite the development and evolution of evaluation to date, there is still a notable absence of boundaries, maturity or clear identity. The findings just explored demonstrate such issues with the identity, purpose and use of program evaluation. The expansive nature of evaluation, and the immense expectations of stakeholders in the complex policy systems it resides, may hamper its ongoing utilisation. For evaluation practitioners, these findings prompt a rethink for how the evaluation community respond to recognition that program evaluation and policy systems are becoming increasingly complex. For a number of reasons (austerity, neo-liberalism) new public governance and notions of co-governance are further complicating the meso-level policy and governance communities. Put simply, more parties are becoming involved in policy concerns and could influence program evaluation. Few tactics to overcome this have been suggested, but network theory and network governance arrangements may be worthy of further exploration, and are purported to be important in complex systems (Walton, 2016). Although, it seems somewhat ironic, given the governance purpose that program evaluation serves, that additional governance is required to govern mechanisms such as evaluation (meta-governance).

The use of metaphorical 'pillars' (of responsibility and governance) central to this book, is interesting to apply to this study. Within architecture, features such as pillars (or 'pilotis'; or 'columns') carry significance beyond their initial structural function, and may also carry aesthetic (Sparrow, 2017) or symbolic relevance (Thacker, 2000). The same sentiment can be applied to the functions of program evaluation that emerged through this study. Evaluation as structural, aesthetic, and symbolic pillars can be identified through the findings - a metaphorical depiction that integrates several disparate arguments within the evaluation discourse. The symbolic use of evaluation, noted earlier, suggests program evaluation activity is used by policy implementers to represent, or even misrepresent, performance. Related, evaluation as aesthetic represents the notion of evaluation '*being seen to*' demonstrate particular outputs or outcomes. The structural use of evaluation represents evaluation as being functional and stabilising; for instance, being used to demonstrate accountability and good governance, or inform learning, policy development, and policy decision-making. The three metaphorical pillars of program evaluation as structural, aesthetic, and symbolic introduced, represent a conceptual framework for further development. Such recognition that program evaluation activity is not always needed for structural reasons such as accountability, governance and learning, prompts further consideration for how symbolic and aesthetic expectations for program evaluation are managed.

For the scholarly community, these findings are confirmatory, highlighting the complex evaluation environment and supporting a growing movement towards 'complexity-appropriate evaluation' (CECAN, 2018). Within evaluation, discourse on methodology (paradigm-oriented approaches) have prevailed, but given the increasing acknowledgement of the complex systems in which program evaluation operates then perhaps more problem-inspired and -oriented approaches to evaluation are necessary.

Finally, the study has implications for the twin pillars of responsibility and governance at the heart of this book. It serves a reminder that governance, and the wider concerns for sustainability, are based in complex systems where even those functions set to enhance affairs (such as program evaluation) can in themselves become complex and confused.

## References

Adelman, C. (1996) Anything goes: Evaluation and relativism, *Evaluation*, 2(3), pp.291-305

Berk, R.A. and Rossi, P.H. (1990) *Thinking about Program Evaluation*, Newbury Park, CA: Sage Publications

Ali, M.A. (2017) Stakeholder Salience for Stakeholder Firms: An Attempt to Reframe an Important Heuristic Device, *Journal of Business Ethics*, 144(1), pp.153-168

Barbier, J. (1999) Inter-governmental Evaluation: Balancing Stakeholders' Expectations with Enlightenment Objectives, *Evaluation*, 5(4), pp.373-385

Bechar, S. and Mero-Jaffe, I. (2014) Who is Afraid of Evaluation? Ethics in Evaluation Research as a Way to Cope with Excessive Evaluation Anxiety: Insights from a Case Study. *American Journal of Evaluation*, 35(3): 365-376

Berk, R.A. and Rossi, P.H. (1990) *Thinking about Program Evaluation*, Newbury Park, CA: Sage Publications

Bevan, M.T. (2014) A Method of Phenomenological Interviewing, *Qualitative Health Research*, 24(1), pp.136-144

Boaz, A. and Nutley, S.M. (2003) 'Evidence-based Policy and Practice' in T. Bovaird and E. Loeffler (eds) *Public Management and Governance*, London: Taylor and Francis

Bovaird, T. and Davis, P. (1996) *'Managing on limited resources: a review of the literature'*, Aston Business School: Birmingham

Bovaird, T. and Loeffler, E. (2007) Assessing the Quality of Local Governance: A Case Study of Public Services, *Public Money & Management*, September 2007, pp. 293-299

Brinkerhoff, D.W., Brinkerhoff, J.M. (2015) Public Sector Management Reform in Developing Countries: Perspectives beyond NPM Orthodoxy, *Public Administration and Development*, 35, pp.222-237

Brinkmann, S (2013). *Qualitative Interviewing*. New York: Oxford University Press

Bristow, D., Carter, L. & Martin, S. (2015) Using evidence to improve policy and practice: the UK What Works Centres, *Contemporary Social Science*, 10(2), pp.126-137

Castellani, B. & Hafferty, F. (2009) Sociology and complexity science: A new field of inquiry, Berlin: Springer

CECAN – Centre for the Evaluation of Complexity across the Nexus (2018), Policy evaluation for a complex world, Manifesto, January 2018, https://www.cecan.ac.uk/sites/default/files/2018-01/CECAN%20Policy%20Evaluation%20for%20a%20Complex%20World%281%29.pdf (Accessed 10th March 2018).

Chelimsky, E. and Shaelish, W.R. (1997) *Evaluation for the 21st Century: A Handbook*, London: Sage Publications

Chelimsky, E. (2008) 'A Clash of Culture: Improving the 'Fit' Between Evaluative Independence and the Political Requirements of a Democratic Society', *American Journal of Evaluation*, 29(4), December 2008, pp.400-415

Chen, H.T. (2014) *Practical Program Evaluation: Theory-driven evaluation and the integrated evaluation perspective*. 2 Ed. Thousand Oaks, CA: Sage Publications

Christie, C.A. and Fleischer, D. (2009) 'Social Inquiry Paradigms as a Frame for the Debate on Credible Evidence' in Donaldson, S.I., Christie, C.A. and Mark, M.M. (2009) *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* London: Sage Publications

Clarke, J. (2004) 'Dissolving the public realm? The logics and limits of neon-liberalism', *Journal of Social Policy*, 33(1), pp.27-48

Crowther, D, Seifi, S. and Moyeen, A. (2017) Responsibility and Governance in Achieving Sustainability, in Crowther, D, Seifi, S., Moyeen, A *The Goals of Sustainable Development: Responsibility and Governance*, Springer: Singapore

Davies, I.C. (1999) 'Evaluation and Performance Management in Government', *Evaluation*, 5(2), pp.150-159

Diamond, J. (2005) 'Reflecting on the Processes of a Local Evaluation: Networks, narratives and partnerships', *International Journal of Public Sector Management*, 18(2), pp.178-189

Donaldson, S.I., Christie, C.A. and Mark, M.M. (2009) *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* London: Sage Publications

Donaldson, S. I., Gooler, L. E., and Scriven, M. (2002). Evaluation strategies for managing evaluation anxiety: Toward a psychology of program. *American Journal of Evaluation*, 23, pp.261–273.

Duncan, G. and Chapman, J. (2012) Better public services: Public management and the New Zealand model, *Public Policy*, 7, pp.151-166

Evers, A., and Ewert, B. (2012). Co-Production: Contested Meanings and Challenges for User Organizations. In T. Brandsen, C. Pestoff, and B. Verschuere, New *Public Governance, the Third Sector and Co-Production*, pp.61–78. New York: Routledge

Feinstein, O.N. (2002) 'Use of Evaluation and the Evaluation of their Use', *Evaluation*, 8(4), pp.433-439

Frederickson, H.G. (2005) 'What happened to public administration? Governance, Governance Everywhere', in E. Ferlie, L. Lynn and C. Pollitt (eds) *The Oxford Handbook of Public Management*, Oxford: Oxford University Press

Fitzpatrick, J., Christie, C. and Mark, M.M. (2009) *Evaluation in Action*, London: Sage Publications Limited

Giannone, D. (2016) Neoliberalization by evaluation: Explaining the making of neoliberal evaluative state, PARTECIPAZIONE e CONFLITTO: The Open Journal of Sociopolitical Studies

Gilbert, N. (2017) Evaluating complexity, *Evaluator*, spring 2017, pp.5-7

Glendinning, C., Powell, M. and Rummery, K. (2002) *Partnerships, New Labour and the Governance of Welfare*, Bristol: Policy Press

Hanberger, A. (2006) 'Evaluation of and for Democracy', *Evaluation*, 12(1), pp.17-37

Hansson, F. (2006) 'Organizational Use of Evaluations: Governance and Control in Research Evaluation', *Evaluation*, 12(2), pp.159-178

Haynes, P. (2008) Complexity Theory and Evaluation in Public Management, *Public Management Review*, 19(3), pp.401-419

Head, B.W. (2008b), Three lenses of evidence-based policy, *Australian Journal of Public Administration*, 67(1), pp. 1-11

Henkel, M. (1991) The New Evaluative State, *Public Administration*, 69, pp.121-136

Huebner, A.J. and Betts, S.C. (1999) Examining fourth generation evaluation: application to positive youth development, *Evaluation*, 5(3), pp.340-58

Husbands, C. (2007) Evaluating Complex Public Policy Programmes: Reflections on Evaluation and Governance from the Evaluation of Children's Trusts, Dilemmas of Engagement, *Evaluation and the New Public Management*, *Advances in Program Evaluation*, 10, pp.53-65

Husserl, E. (1970) *The crisis of European sciences and transcendental phenomenologies: An introduction to phenomenological philosophy*, Evanston, IL: Northwestern University Press

Iriti, J.E., Bickel, W.E. and Nelson, C.A. (2005) Using Recommendations in Evaluation: A decision-making framework for evaluator, *American Journal of Evaluation*, 26(4), pp.464-479

Jones, B., Kavanagh, D., Morgan, M. and Norton, P. (2007) *Politics UK*, 6th Ed. Pearson Education Limited: Essex
Kettl, D.F. (2005) *The Global Public Management Revolution*, 2nd Ed. Washington, DC: Brookings Institute
Klijn, E-H. (2008) Governance and governance networks in Europe, *Public Management Review*, 10(4), pp.505-525
Mason, J. (2002) *Qualitative researching*, 2nd Ed. London: Sage Publications
McCoy, M. and Hargie, O.D.W. (2001) 'Evaluating Evaluation: Implications for Assessing Quality', *International Journal of Health Care Quality Assurance*, 14(7), pp.317-327
McKie, L. (2003) Rhetorical Spaces: Participation and Pragmatism in the Evaluation of Community Health Work, *Evaluation*, 9(3), pp.303-323
Meek, J.W. (2014) Complexity theory and administrative learning – adaptive practices in complex governance systems, *Emergence: Complexity and Organization*, Editorial, 16(1), pp.1-6
Mohan, R. and Sullivan, K. (2006) 'Managing the Politics of Evaluation to Achieve Impact', *New Directions for Evaluation*, winter 2006, 112, pp.7-23
Morrell, J.A. (2010) Evaluation in the face of uncertainty, anticipating surprise and responding to the inevitable, New York, NY: Guildford Press
National Audit Office (2013) Evaluation in Government, Report, https://www.nao.org.uk/wp-content/uploads/2013/12/10331-001-Evaluation-in-government_NEW.pdf (Accessed 10th March 2018).
Newcomer, K. (1997) 'Using Performance measurement to improve public and non-profit programs', *New Directions for Program Evaluation*, 75, San Francisco, CA: Jossey-Bass
Neville, B.A., Bell, S.J. and Whitwell, G.J. (2011) Stakeholder salience revisited: Refining, redefining, and refuelling an underdeveloped conceptual tool, *Journal of Business Ethics*, 102, 357-378
O'Brien, T. Payne, S. Nolan, M and Ingleton, C. (2010) Unpacking the Politics of Evaluation: A Dramaturgical Analysis, *Evaluation*, 16(4), p.431-444
Osborne, S. (2000) Public-private partnerships: Theory and Practice in International Perspective, London: Routledge
Parkhurst, J. (2017) The politics of evidence: from evidence-based policy to the good governance of evidence. Routledge Studies in Governance and Public Policy. Routledge, Abingdon, Oxon, UK
Patton, M.Q. (1997) *Utilization-focused Evaluation: The New Century Text*, 3rd Ed. London: Sage Publications
Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*, London: Sage Publications
Picciato R (1999) Towards an economics of evaluation. *Evaluation* 5(1), pp.7-22
Plottu, B. and Plottu, E. (2009) Approaches to Participation in Evaluation: Some Conditions for Implementation. *Evaluation*, 15 (3), pp.343-359
Pollitt, C. (2003) *The Essential Public Manager*, Open University Press: Philadelphia
Roe, E. (1982) Reflections of an Evaluator, *Higher Education Research & Development*, 1(2), pp.129-142
Rossi, P.H., Lipsey, M.W. and Freeman, H.E. (2003) *Evaluation: A Systematic Approach*, 7th Ed., Newbury Park, CA: Sage Publications
Sanderson, I. (2000) Evaluation in Complex Policy Systems, Evaluation, 6(4), pp.433-454
Scriven, M. (1996) The theory behind practical evaluation, *Evaluation*, 2(4), pp.393-404
Scriven, M. (1980) *The Logic of Evaluation*, Inverness, CA: Edgepress

Seppanen-Jarbela, R. (2003) Internal Evaluation of a Management-development Initiative: A Public Sector Case, *Journal of Management Development*, 24(1), pp.45-56
Sparrow, J.G. (2017) *Modern Jordan*, Routledge, Abingdon, UK
Shaw, I. (1999) *Qualitative Evaluation*, London: Sage Publications
Shaw, I. and Faulkner, A. (2006) Practitioner Evaluation at Work, *American Journal of Evaluation*, 27(1), pp.44-63
Stecher, B.M. and Davis, W.A. (1988) *How to Focus an Evaluation*, 2nd Ed. London: Sage Publications
Stern, E. (2008) Evaluation: Critical for Whom and Connected to What? *Evaluation*, 14(2), pp.249-257
Suchman, M.C. (1995) Managing Legitimacy: Strategic and Institutional Approaches, *The Academy of Management Review*, 20(3), pp. 571-610
Taylor, D. (2005) Governing through Evidence: Participation and Power in Policy Evaluation, *Journal of Social Policy*, 34(4), pp.601-618
Thacker, A. (2000) Toppling Masonry and Textual Space: Nelson's Pillar and Spatial Politics in Ulysses, Irish Studies Review, 8(2), pp.195-203
Theisens, H., Hooge, E. and Waslander, S. (2016). Steering dynamics in complex education systems. An agenda for empirical research. *European Journal of Education*, 51(4), 463–477
Tompa, E., Culyer, A. J. and Dolinschi, R. (2008). Economic Evaluation of Interventions for Occupational Health and Safety: Developing Good Practice. Oxford: Oxford University Press
Tourmen, C. (2009) 'Evaluators' Decision Making: The Relationship between Theory, Practice, and Experience', *American Journal of Evaluation*, 30(7), pp.7-30
Tufford, L. and Newman, P. (2010) Bracketing in Qualitative Research, *Qualitative Social Work*, 11(1), pp.80-96
Walton, M. (2016) Setting the context for using complexity theory in evaluation: boundaries, governance and utilisation, Evidence & Policy, 12(1), pp.73-89
Weiss, C.H. (1972) *Evaluation Research*, Prentice Hall, Englewood Cliffs: NJ
Weiss (1993) Where Politics and Evaluation Research Meet, *Evaluation Practice*, 14(1), pp.93-106
Wond, T. (2017) Trust matters: Distrust in an external evaluation of a public sector program, *International Journal of Public Administration*, 40(5), pp. 408-415
Wond, T. and Macaulay, M. (2010) 'Evaluating local implementation: An evidence- based approach', *Policy and Society*, 29(2), pp.161-169



## Tables, Figures and Diagrams

| Evidence for Transparency | Evidence for Accountability |
|---|---|
| • To demonstrate where public funds are spent (Glendinning et al, 2002; Clarke, 2004; Jones et al, 2007). | • To demonstrate that public funds have been spent well and provided value for money (Barbier, 1999; Davies, 1999; Huebner and Betts, 1999; McCoy and Hargie 2001).<br>• To substantiate results, outcome, impact, and outputs (Berk and Rossi, 1990; Newcomer, 1997; Sanderson, 2000). |
| Evidence for Responsibility | Evidence for Fairness |

| | |
|---|---|
| • To demonstrate that public funds have been spend responsibly.<br>• To develop/improve/learn in order to more effectively allocate future spend (Stecher and Davis, 1988; Patton, 1997; Shaw, 1999; Shaw and Faulkner, 2006). | • To demonstrate that processes deemed fair have been followed in spending public funds.<br>• To demonstrate distributive and procedural equity in who came to benefit from public funds (Tompa et al, 2008) |

*Table 1: Evidence for Governance - Source: Author's own*