Evaluation for What Purpose? Findings From Two Stakeholder Groups

Tracey Wond

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 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 inherent complexity within them, may lead to conflicting expectations about why 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 evaluators and policy implementers towards the purpose of evidence. The findings suggest evaluators and policy implementers have divergent expectations of why and how evaluation data might be used. The findings suggest that evaluators aspire to make a change and enhance the policy domains they serve, whereas policy implementers perceive evaluation as serving a more governance-/management-orientated role. The use of evaluation as a symbolic or structural mechanism also emerges, prompting opportunity for further research, for instance, to explore legitimacy and evaluation. The chapter demonstrates the complexity of both 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.

Keywords Evaluation utilisation · Governance · Evidence-based policy and practice · Complexity

T. Wond (✉)
University of Derby, Derby, UK
e-mail: T.Wond@derby.ac.uk

1   Introduction

The growth of evaluation has been heavily linked to public reform and the emergence of the New Public Management (NPM) ideology (Shaw 1999). Through NPM, public sectors across the world have engaged heavily with evidential mechanisms such as evaluation (Taylor 2005). Evaluation provides a means to ensure good governance, assuring value for money, efficiency, and accountability (Boaz and Nutley 2003; Bovaird and Loeffler 2007; Stern 2008).

Tension over the purpose of evidence and role of evaluators has mounted over several decades. Adelman (1996, p. 295) acknowledges the tense and contradictory relationship between evaluation and policymakers: 'evaluators want to influence policy-making, but few were willing to participate in the process of decision-making; that was the responsibility of policymakers'. Others have noted 'growing disagreement and confusion about what constitutes sound evidence for decision-making' (Donaldson et al. 2009, p. 12). The extensiveness of the evaluation concept can be problematic and confusing (Weiss 1972; McKie 2003), and this has heightened as evaluation has evolved with sophisticated methodologies emerging to demonstrate responsible public spending, demonstrate success and identify what works (Bristow et al. 2015). This chapter, and the research underpinning it, responds to recognition that a dearth of studies explores the practice of evaluation (Fitzpatrick et al. 2009). A decade earlier, Pawson and Tilley (1997, p. 24) suggested 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 much 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.

Further, there is increasing acknowledgement that public policy, governance arrangements and evaluation coexist in complex environments (Walton 2016). It is reasonable to assume that as models of public administration become more polycentric and more stakeholders become involved in policy development, decision-making and governance, this complexity will intensify (Evers and Ewert 2012). This chapter acknowledges these complex systems.

The study on which this chapter is based sought to identify and compare the perceptions of evaluators and policy implementers towards the purpose and use of evidence. It is based on semi-structured interviews with 19 evaluators (internal and external) and 10 policy implementers (managers and senior personnel of publicly funded programmes) in the UK. The study, therefore, explores evidence used 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 (Berk and Rossi 1990; Newcomer 1997; Davies 1999; Taylor 2005; Hansson 2006; Bovaird and Loeffler 2007; Stern 2008). 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 evaluation and policy concerns such as environmental sustainability reside.

2 Literature Review

2.1 *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). Evaluation extends several disciplines, attracting the label of a 'metadiscipline' (Picciato 1999, p. 7) and 'transdiscipline' (Scriven 1996, p. 402). These labels provide an example of the breadth of the evaluation definition and the scope for evaluation to be misunderstood by the multiple stakeholders engaged with it. 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'. The extensiveness of the evaluation concept results in several varying purposes of evaluation emerging and there have been some attempts to characterise these in the literature (Husbands 2007; Shaw and Faulkner 2006; Berk and Rossi 1990).

2.2 *Purposes of Evaluation*

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).

2.2.1 Evaluation for Governance

Governance has become a catchall 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 the following four main definitions of governance:

- Governance as 'good/corporate governance', relating to the fair and proper operation of government;
- Governance as 'new public management', including embedding performance measurement and accountability mechanisms for those delivering public services;
- Governance as 'multi-level' and/or 'inter-governmental', embracing the multiple layers and hierarchies of public organisations;
- Governance as 'network governance', which implies the need to manage across networks and complex systems, that often cross boundaries are involve multiple actors.

Emerging from these four definitions of governance, it appears that governance extends beyond denoting an activity (i.e. something that is done), in a single organisation (ensuring fairness, evaluating). Governance by Klijn's (2008) definitions appears more pervasive, difficult to bound and scalable. This is typical of a complex system (CECAN 2018).

Crowther et al. (2017) identify good governance as comprising four parts, '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.

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 and 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).

Table 1 Evidence for governance

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). • 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 spent responsibly. • 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. • To demonstrate distributive and procedural equity in who came to benefit from public funds (Tompa et al. 2008)

Source Author's own

Attempts by public services to be more responsible, strategic and efficiency 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 neoliberalism (Giannone 2016). Greater public voice and media scrutiny have also pressured public services to demonstrate responsible stewardship of 'taxpayers' money' (Barbier 1999, p. 378).

2.2.2 Evaluation for Learning

The potential for evaluation to serve a learning or developmental purpose are also well acknowledged (Weiss 1972; McCoy and Hargie 2001; McKie 2003; Shaw and Faulkner 2006; Fitzpatrick et al. 2009). This includes development programme 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. Scriven's (1980, p. 6) formative/ summative dualism identifies the provision of feedback 'to improve something', and the provision of knowledge to decision-makers as central to formative evaluation.

Many have contended this learning purpose (Bovaird and Davis 1996; Iriti et al. 2005), in particular, questioning the impartiality 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 'downstream', towards development and implementation is not necessarily a new role for evaluation and resonates with its historical roots in supporting US reform. However, it does represent scope for confusion and conflicting expectations (Weiss 1972; McKie 2003; Donaldson et al. 2009).

2.3 Challenges Facing Evaluation and Evidence Use

The under-utilisation and effectiveness of evaluation findings have 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 Brexit impact assessments, to be disclosed in a timely manner are examples of this.

The changing face of the public sector and form that governance takes may also have implications on 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 suggest 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).

A host of elaborate methodologies and perspectives of what makes good evidence have emerged (reigniting the quantitative/qualitative paradigm war in doing so), and have further alienated policymakers from the evidence bases meant to help them (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.

2.4   Complexity Theory

Whilst this study does not seek to explore complexity theory in detail, nor make it a key feature of this chapter, the theory does acknowledge the complex systems underpinning the setting for this study, and is therefore 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 evaluation settings, both as a methodological to evaluate and to study evaluation (Haynes 2008; Morrell 2010; Walton 2016; Gilbert 2017), and there is increasing awareness that the complex policy systems in which evaluation operates require such an approach (Walton 2016; CECAN 2018). 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).

3 Methodology

The study explored in this chapter provides a meso-level exploration of two types of communities (evaluators and policy implementers) participating in 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'. Both policy implementers and 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 practising evaluators (9 female and 10 male) who undertake evaluation in various capacities (academics, evaluation consultants, internal evaluators undertaking programme evaluation). Interviews with 10 policy implementers were also conducted. Both groups represented a range of policy areas (e.g. health, education, foreign aid and enterprise support). The interviews were undertaken as part of a wider study and looked to understand respondents wider experiences of evaluation, challenges they felt limited evidence use, and how they used or perceived the use of evaluation evidence. Interviews were administered via telephone, face-to-face and Skype. A subset of this data, which related to the purpose and perceived use of evaluation is analysed here.

The researcher's own involvement with evaluation societies and networks in the UK and Europe supported access to evaluators. Further networking and involvement in project settings led the researcher to access policy implementers for this research. A semi-structured, informal interviewing approach, without a strict interview guide (Brinkmann 2013), was adopted. Interviews were transcribed contemporaneously in the most part, although in the case of some face-to-face exchanges these were undertaken retrospectively with the use of paraphrasing. Data was analysed using NVIVO.

The findings presented below are abridged, since the wider study focused on many more aspects of evaluation perceptions and practice.

4 Findings

4.1 Making the World Better?

Many evaluators emphasised that they hoped their work would be used to '*make a difference*' and there was a strong association with 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*'. Therefore, a strong moral purpose was evident amongst evaluators. In contrast, only two of the ten policy implementers referenced the potential for evaluation to be used for improvement ('*feed(ing) the evidence-base*' and '*showing us what works*'). There was clear disparity between how evaluators hoped their work would be used, and how policy implementers perceived the use of evaluation.

4.2 Evaluation for Governance

The majority of policy implementers perceived that evaluation should play a heavy governance and monitoring role, and they spoke of evaluation as instrumental in proving targets and assuring responsible spending, for instance: '*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*'. Evaluators also recognised this governance role and the pursuit of '*the usual monitoring 'stuff' (data)*' emerged in 11 of the 19 interviews (and more so amongst internal evaluators).

4.3 Product of Evaluation

There was recognition by both groups that evaluation reports were under-utilised: Evaluators spoke of evaluation reports '*gathering dust*', '*abandoned*' and '*lost*' in office drawers, and '*unread in an inbox somewhere*'. Policy implementers spoke less of 'evaluation reports', as a product of 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*') and demonstrate targets had been met within particular reporting periods.

4.4 Evaluation as Symbolic

Evaluation was frequently spoken of as supporting programmes/policy interventions to be '*seen to*' deliver, 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 evaluator also recognised that evaluation was perceived in this way). As such, a symbolic role for evaluation was also recognisable.

5 Discussion

There are several implications from the findings presented, particularly given the contradictory perceptions of the use of evaluation.

5.1 Divergence in Perceptions

An incongruence between the supply of evaluation (by evaluators) and demand for 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 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 identification and salience, and complexity literature. Stakeholder theories acknowledge that cooperation and collective action supports the salience of particular stakeholders (Ali 2017). Policy implementers and evaluators in this study appeared to differ on matters such as the use of evaluation reports, and the purpose of evaluation as a whole. 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). Walton (2016) suggests that network governance arrangements could be introduced to address conflicts in complex systems, although it seems somewhat ironic given the governance purpose that evaluation serves that additional governance is required to govern it (meta-governance).

The moral mission of evaluators, seen in the findings, prompts discussion about the motivation of evaluators to fulfil a moral or social mission and the implications of this. Evaluators overwhelmingly felt they served to, in the words of one respondent, '*make the world a better place*'. This finding associates evaluation with utility, and ultimately the end-users and the policy implications of evaluation evidence. This is despite much literature and acknowledgement in this study (by both policy implementers and evaluators) that evaluation is often under-utilised; this moral position could, therefore, be considered ideological. Further study to explore the motivations of evaluators could be valuable.

5.2 Implications for Governance

Evaluation for accountability and transparency emerged through many of the interviews and a clear governance role for 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 referred to the basic monitoring function underpinning evaluation, as opposed to more elaborate or technical modes of evaluation. The evaluation community may be disappointed to note such findings, since such perceptions (by policy implementers) appear to oversimplify 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' 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 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.

5.3 Symbolic Versus Structural Use of Evidence

The metaphorical use of 'pillars' (of responsibility and governance) central to this book are particularly relevant to the discovery that evaluation appeared to be treated symbolically, almost aesthetically—a tick-box exercise. Within architecture, features such as pillars ('pilotis','columns') carry significance beyond their initial structural function, and may also carry aesthetic (Sparrow 2017) or symbolic relevance (Thacker 2000). The same appeared true in this study.

Reference was made to evaluation making programmes and policy interventions 'be seen' in a particular way (successful, meeting targets, etc.). 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. Much attention has been paid to strategic legitimacy in respect of social or environmental reporting for instance.

The symbolic use of evaluation sits in contrast with the functional and structural potential for evaluation evidence to be used to inform learning, policy development and policy decision-making. Such challenges in the pseudo/symbolic use of evidence for governance exposes potential vulnerabilities in the triadic relationship between sustainability, governance and responsibility at the heart of this book.

6 Conclusions

Despite evaluation existing for many years and evidence of its evolution, 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 evaluation. The expansive nature of evaluation, and the immense expectations of stakeholders in the complex policy systems it resides may hamper its utilisation.

These study findings have implications for practitioner and scholarly communities. For evaluation practitioners, they prompt a rethink for how evaluators and the evaluation function respond to recognition that 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 evaluation. Few tactics to overcome this have been suggested, but network theory and network governance arrangements may be worthy of further exploration (Walton 2016). The motivations of evaluators to fulfil a 'moral mission' emerged as an interesting finding worthy of further exploration. However, 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. For the scholarly community, these findings are in many respects confirmatory, continuing to link legitimacy and evaluation, and complexity and evaluation.

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 evaluation) can in themselves become complex and confused.

References

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

Ali, M. A. (2017). Stakeholder salience for stakeholder firms: An attempt to reframe an important heuristic device. *Journal of Business Ethics, 144*(1), 153–168.

Berk, R. A., & Rossi, P. H. (1990). *Thinking about program evaluation*. Newbury Park, CA: Sage Publications.

Barbier, J. (1999). Inter-governmental evaluation: Balancing stakeholders' expectations with enlightenment objectives. *Evaluation, 5*(4), 373–385.

Boaz, A., & Nutley, S. M. (2003). Evidence-based policy and practice. In T. Bovaird & E. Loeffler (Eds.), *Public management and governance*. London: Taylor and Francis.

Bovaird, T., & Davis, P. (1996). *Managing on limited resources: A review of the literature*. Aston Business School: Birmingham.

Bovaird, T., & Loeffler, E. (2007). Assessing the quality of local governance: A case study of public services. *Public Money & Management*, September 2007, pp. 293–299.

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), 126–137. https://doi.org/10.1080/21582041.2015.1061688.

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.

Chelimsky, E., & Shaelish, W. R. (1997). *Evaluation for the 21st century: A handbook*. London: Sage Publications.

Christie, C. A., & Fleischer, D. (2009). Social inquiry paradigms as a frame for the debate on credible evidence. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *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), 27–48.

Crowther, D., Seifi, S., & Moyeen, A. (2017). Responsibility and governance in achieving sustainability. In D. Crowther, S. Seifi, & A. Moyeen (Eds.), *The goals of sustainable development*. Responsibility and Governance, Springer: Singapore.

Davies, I. C. (1999). Evaluation and performance management in government. *Evaluation, 5*(2), 150–159.

Donaldson, S. I., Christie, C. A., & Mark, M. M. (2009). *What counts as credible evidence in applied research and evaluation practice?*. London: Sage Publications.

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

Evers, A., & Ewert, B. (2012). Co-production: contested meanings and challenges for user organizations. In T. Brandsen, C. Pestoff, & B. Verschuere (Eds.), *New public governance, the third sector and co-production* (pp. 61–78). New York: Routledge.

Frederickson, H. G. (2005). What happened to public administration? Governance, Governance Everywhere. In E. Ferlie, L. Lynn, & C. Pollitt (Eds.), *The oxford handbook of public management*. Oxford: Oxford University Press.

Fitzpatrick, J., Christie, C., & 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,* 5–7.

Glendinning, C., Powell, M., & Rummery, K. (2002). *Partnerships, new labour and the governance of welfare.* Bristol: Policy Press.

Hansson, F. (2006). Organizational use of evaluations: governance and control in research evaluation. *Evaluation, 12*(2), 159–178.

Haynes, P. (2008). Complexity theory and evaluation in public management. *Public Management Review, 19*(3), 401–419.

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

Henkel, M. (1991). The new evaluative state. *Public Administration, 69,* 121–136.

Huebner, A. J., & Betts, S. C. (1999). Examining fourth generation evaluation: Application to positive youth development. *Evaluation, 5*(3), 340–358.

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,* 53–65.

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

Jones, B., Kavanagh, D., Morgan, M., & Norton, P. (2007). *Politics UK* (6th ed.). Pearson Education Limited: Essex.

Kettl (2005). *British Politics: A beginners guide*. Oxford: Oneworld Publications.

Klijn, E.-H. (2008). Governance and governance networks in Europe. *Public Management Review, 10*(4), 505–525.

McCoy, M., & Hargie, O. D. W. (2001). Evaluating evaluation: Implications for assessing quality. *International Journal of Health Care Quality Assurance, 14*(7), 317–327.

McKie, L. (2003). Rhetorical spaces: Participation and pragmatism in the evaluation of community health work. *Evaluation, 9*(3), 303–323.

Meek, J. W. (2014). Complexity theory and administrative learning—Adaptive practices in complex governance systems. *Emergence: Complexity and Organization Editorial, 16*(1), 1–6.

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.

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., & Whitwell, G. J. (2011). Stakeholder salience revisited: Refining, redefining, and refuelling an underdeveloped conceptual tool. *Journal of Business Ethics, 102*, 357–378.

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., & Tilley, N. (1997). *Realistic evaluation*. London: Sage Publications.

Picciato, R. (1999). Towards an economics of evaluation. *Evaluation, 5*(1), 7–22.

Plottu, B., & Plottu, E. (2009). Approaches to participation in evaluation: Some conditions for implementation. *Evaluation, 15*(3), 343–359.

Sanderson, I. (2000). Evaluation in complex policy systems. *Evaluation, 6*(4), 433–454.

Scriven, M. (1996). The theory behind practical evaluation. *Evaluation, 2*(4), 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), 45–56.

Sparrow, J. G. (2017). *Modern Jordan*. Abingdo, UK: Routledge.

Shaw, I. (1999). *Qualitative evaluation*. London: Sage Publications.

Shaw, I., & Faulkner, A. (2006). Practitioner evaluation at work. *American Journal of Evaluation, 27*(1), 44–63.

Stecher, B. M., & 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), 249–257.

Taylor, D. (2005). Governing through evidence: Participation and power in policy evaluation. *Journal of Social Policy, 34*(4), 601–618.

Thacker, A. (2000). Toppling masonry and textual space: nelson's pillar and spatial politics in ulysses. *Irish Studies Review, 8*(2), 195–203. https://doi.org/10.1080/713674244.

Theisens, H., Hooge, E., & 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., & Dolinschi, R. (2008). *Economic evaluation of interventions for occupational health and safety: Developing good practice*. Oxford: Oxford University Press.

Walton, M. (2016). Setting the context for using complexity theory in evaluation: boundaries, governance and utilisation. *Evidence & Policy, 12*(1), 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), 93–106.

Wond, T. (2017). Trust matters: Distrust in an external evaluation of a public sector program. *International Journal of Public Administration, 40*(5), 408–415.