Methods for Delivering Complex Social Services: Exploring adaptive management and regulation in the australian national disability insurance scheme

Gemma Carey & Mark Matthews

⊞ Published online: 11 Mar 2016.

�✎ Submit your article to this journal ⟐

◉Q View related articles ⟐

CrossMark View Crossmark data ⟐

Abstract

In the wake of new public management, and with the emergence of new public governance, a number of challenges remain unsolved in the field of public service governance and management. In this paper, we use the example of the Australian National Disability Insurance Scheme to show how we might deal with the design and implementation of a public service reform within the context of new public governance. We argue that governments need to develop greater openness to risk and policy experimentation during implementation. We therefore propose an adaptive system architecture that could support such risk taking.

METHODS FOR DELIVERING COMPLEX SOCIAL SERVICES

Exploring adaptive management and regulation in the australian national disability insurance scheme

Gemma Carey and Mark Matthews

Gemma Carey
Regulatory Institutions Network
Australian National University,
Canberra,
Australia
Email: Gemma.carey@anu.edu.au

Mark Matthews
Crawford School for Public Policy
Australian National University
Canberra
Australia
Email: mark.matthews@me.com

Routledge
Taylor & Francis Group

INTRODUCTION

In the wake of new public management, and the emergence of new public governance, a number of challenges remain unsolved in the field of public service governance and management (Osborne 2010, 2006). While influenced by new public management, new public governance draws strongly on network and institutional theoretical perspectives with the aim of capturing the real world complexity of the design, implementation and management of public policy in (Kickert and Koppenjan 1997; Osborne 2010; Rhodes 1997). In doing so, it has been suggested that new public governance pushes us beyond new public management, which has 'been criticized most devastatingly for its intra-governmental focus in an increasingly plural world and for its adherence to the application of outdated private-sector techniques' (Osborne 2006, 308).

New public governance perspectives posit a *plural* state, with multiple interdependent actors contributing to the delivery of services, as well as a *pluralist* state, where multiple processes inform the policy-making system (Osborne 2010). These two aspects of plurality mean the focus within and new public governance approach is on inter-organizational relationships and governance and regulatory processes.

In this paper, we use the example of the Australian National Disability Insurance Scheme (NDIS), as an example of new public governance in action, to flesh out how we might deal with these multiple forms of plurality in the design and implementation of a major public service reform. The NDIS is a useful representative example to consider because it confronts two of the most problematic issues in new public governance: the use of quasi-market arrangements in large-scale public service delivery, and how to ensure that flexible and effective adaptive governance and regulation processes (and/or architectures) are in place (Braithwaite 2008; Carey and Crammond 2015b; Osborne 2010). As such, the NDIS is a notable expression of the more general 'new public governance' ethos of developing complex pluralistic policy delivery mechanisms.

We draw on the literature on experimentation in policy and adaptive management to propose a Bayesian model for conceptualizing how learning and adaptation can be fostered and institutionalized with regard to social policy reforms (Sabel 1995; Williams et al. 2012). The proposed approach also has the advantage that it provides a basis for assessing and demonstrating the value-for-money of experimental and adaptive modes of governance – a dimension currently missing from established approaches of this type.[1] Such thinking, we argue, can help overcome the types of implementation gaps that have plagued many areas of social policy – and show no sign of abating as we shift into new public governance paradigms (Barrett 2004; Carey, McLoughlin, and Crammond 2015; Hill and Hupe 2009).

THE AUSTRALIAN NATIONAL DISABILITY INSURANCE SCHEME

Described as a 'once in a generation' reform, the NDIS is the most significant change to Australia's social protection policies since the creation of the Medicare programme in

the 1970s. Passed in 2013 with broad public and political support, the NDIS will provide no-fault insurance cover for Australians who are born with or acquire a disability and, more recently, includes people experiencing forms of mental illness. It is expected to secure improvements in the lives of hundreds of thousands of Australians living with a disability, their families and their carers. Current estimates put the cost of the NDIS at $15 billion a year (Australian Productivity Commission 2011). The NDIS has been described as a 'monster of a government program', set to grow (through indexation and population growth) to a cost of $29.5 billion per year by 2023, assisting 500,000 people and supported by 8,000 bureaucrats (Baker 2012). The complexity of the scheme and the subsequent investment required to establish and maintain it makes it the largest and most challenging social policy reform of recent times within the Australian context.

Under the NDIS, eligible individuals will be encouraged and supported to exercise choice and control over a needs-based funding envelope to purchase supports that most effectively meet their needs (Bonyhady 2014; KPMG 2014; NDIS 2014a). For individuals to exercise choice there needs to be a mature market in disability support services. However, adequacy of supply across all market segments (e.g. geography, age and service type) cannot be guaranteed without government intervention (Nevile 2013). Facilitating the development of a disability services market that has the capacity to deliver all types of services to all eligible individuals is one of the most challenging aspects of the NDIS – the intended marketplace will be a world first (LeGrand 2007; LeGrand and Bartlett 1993).

The 2014 interim report into the NDIS report noted that further evidence is needed on what drives different market outcomes in different locations and market segments (KPMG 2014; see also NDIS 2015). The NDIS is said to require a 'detailed market architecture' which clarifies the role of government in creating and intervening in the newly established disability service market (KPMG 2014). A critical component of this is the development of a set of 'intervention levers', which will shape and guide the market (KPMG 2014). To be effective, the market architecture needs to be integrated with institutional governance arrangements in order to be responsive to arising regulatory and management issues (Considine 2003; Larkin and Dickinson 2011; Nevile 2013).

At present, governance and market architectures are being developed through trial sites (NDIS 2015). Both during, and after, trial the NDIS relies upon a continual process of change. The market (and associated regulatory mechanisms) will require continuous monitoring and adjustment (LeGrand 2007), while cross-boundary working is an evolving process requiring actors to be flexible and adaptive to different contexts, shifting norms and values and unfolding institutional change (Carey, McLoughlin, and Crammond 2015; Carey and Crammond 2015b; O'Flynn 2013; Wong et al. 2013, 2012). An important question for administration of complex, market-based approaches to social protection, such as the NDIS, is whether information (from evaluations, reporting and monitoring) can be fed back and whether dual governance and market architectures can be adjusted accordingly.

This aspect is particularly important because one benefit of quasi-market-based approaches is said to be that they encourage and facilitate innovation. This means that unexpected developments that have not been anticipated but have emerged from on-the-ground practices and experimentation can be welcome. However, it also means that negative unexpected consequences are also possible. Consequently, evaluation methods must be 'developmental' in the sense that they allow for evolution rather than framing target activities, outcomes and impacts exclusively on the basis of original concepts and expectations (Hallsworth 2011; Patton 2010).

Several interim reports into the implementation of the NDIS identified the need for institutional arrangements that promote accountability, flexibility and control between different levels of government and the National Disability Insurance Agency, which is charged with oversight of the NDIS (Joint Standing Committee on the National Disability Insurance Scheme 2014; KPMG 2014). In interim reports, not surprisingly, the implementation of the NDIS was found to be constrained by governance and accountability arrangements that promote duplication, a one-way flow of information from strategic to operational levels and a general lack of clarity regarding shared responsibilities between commonwealth and state agencies (KPMG 2014). These are perennial challenges in cross-government and cross-sectoral policy implementation and have been found to be key determinants of policy and programme success (Carey and Crammond 2015a). Hence underdeveloped, or unresponsive, governance and regulation arrangements leave the NDIS vulnerable to implementation deficits (Hill and Hupe 2009). Figure 1 provides an overview of the planned governance structures of the NDIS.

The evidence base on cross-boundary policy implementation (i.e. that which requires coordinated action across and between levels of government, non-government and private sectors) indicates that successful implementation requires a sophisticated and flexible supportive architecture (O'Flynn 2013; O'Flynn et al. 2011). Specifically, this architecture must support bottom-up working (as opposed to the current top-down approach), decentralized control to facilitate local responsiveness and flexibility, and an adaptive set of instruments to facilitate necessary changes in management, processes and cultural and institutional norms (6 1997; Carey, McLoughlin, and Crammond 2015; O'Flynn 2013; Pollitt 2003; Williams et al. 2012). It is worth stressing that this bottom-up ethos seeks to counteract the tendency for bureaucracies to resist bottom-up evolutionary change and only succumb to the imperative for widespread change when crisis ('tipping') points are reached – at which point the organization traditionally responds at a systematic level driven by top-down imperatives (Crozier 1964).

Research suggests that getting this 'supportive architecture' right is both critical and one of the most challenging dimensions of implementing a cross-boundary policy initiative (Carey, Crammond, and Riley 2015; O'Flynn 2013). Effective implementation requires sensitivity to the complexity of the task and the normative issues at play (Hill and Hupe 2009). Governance arrangements would therefore be expected to evolve over time, in response to change and unfolding areas of concern. To this end, accountability, flexibility and control structures have been recommended for continued monitoring and refinement (KPMG 2014; NDIS 2014b).
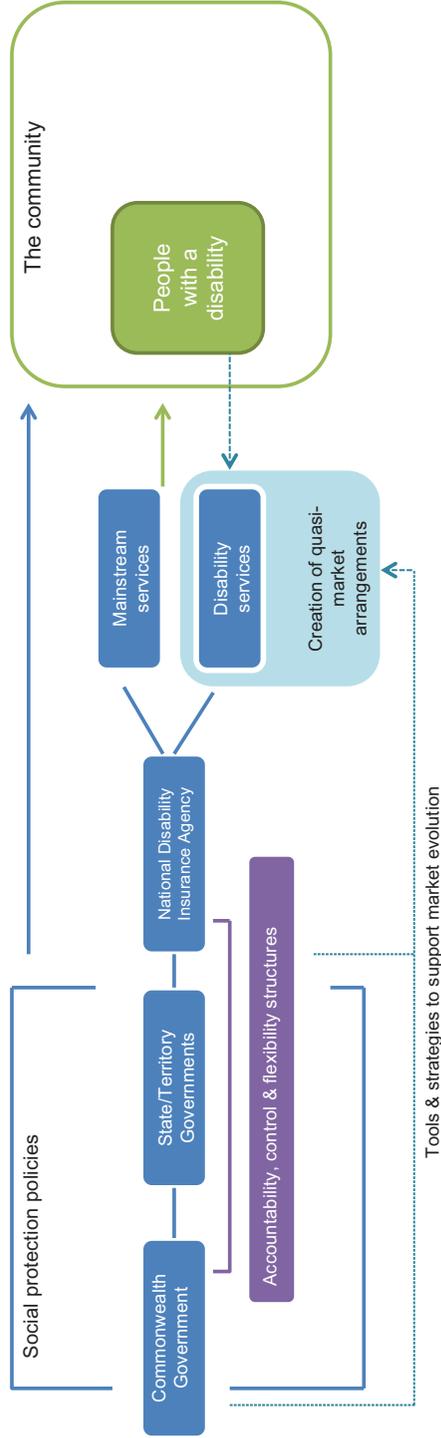
The community

People with a disability

Mainstream services

Disability services

Creation of quasi-market arrangements

National Disability Insurance Agency

Accountability, control & flexibility structures

State/Territory Governments

Commonwealth Government

Social protection policies

Tools & strategies to support market evolution

Figure 1: Governance structures of the NDIS

THE NEED FOR LEARNING AND ADAPTATION IN SOCIAL SERVICES REFORMS

It has been argued that experimentation and adaptation is likely to be an effective way to tackle policy problems, particularly if they are complex (such as the NDIS) (Dorf and Sabel 1998; Hallsworth 2011; Noonan, Sabel, and Simon 2009; Sabel 1995). Here, both problems and the programmes or services being delivered are too multifaceted to be effectively addressed by top-down policy approaches: 'directive approaches are rarely suitable to dealing with complex problems . . . and ongoing public service reforms mean that the systems through which policies are delivered are likely to become even more complex. These changes suggest that governments should increasingly be in the position of setting high-level, resilient goals, and letting the system find the best solution through adaptation and experimentation' (Hallsworth 2011, p. 13). Indeed, Sable has long argued for experimentalist approaches to policy that builds continuous consideration of the systems' 'norms in the course of evaluating them' (Noonan, Sabel, and Simon 2009, p. 52; Sabel 1995). How these experimental learning and adaptive approaches can be put into practice in the context of major reforms (including their accurate expression in the monitoring and evaluation systems that assess value-for-money) remains a compelling question and important area for investigation.

In the context of the NDIS, we propose two integrated adaptive-learning cycles – one at the level of client identification using Bayesian analysis, which is then embedded in a broader cycle of learning secured through flexible governance arrangements that is compatible with this Bayesian approach.

A SIMPLIFIED BAYESIAN APPROACH TO CLIENT IDENTIFICATION WITHIN SERVICE DELIVERY SYSTEMS

The sequential elimination of competing hypotheses using binary classifications (i.e. whether a given hypothesis is true or false given the available data) is the standard diagnostic technique used by physicians (known as 'differential diagnosis'). This approach allows for a structured approach to highly complex and ambiguous conditions to be carried out.

This established diagnostic method in clinical medicine is now being informed by the adoption of techniques used in signal processing and machine learning that focus on the incidence and clinical implications of incorrect test results (false positives and false negatives). This emphasis on the test accuracy dimension now constitutes an important element in 'evidence based medicine'. This is a welcome development that the growing use of sophisticated diagnostic scanners has helped to stimulate. This is because clinicians using these machines must factor likely false positive and false negative test results into their conclusions (the documentation on many machines provides statistical

data on this aspect of their use). See Matthews and Kompas (2015) for a discussion of these issues in relation to risk management in the public sector.

The high complexity and evolutionary nature of interventions like the NDIS suggests that it may be useful to consider the benefits of adopting these signal processing and machine learning approaches. From a regulatory perspective, these approaches are relevant because the misdiagnosis of performance in complex and evolving situations generates obvious risks for policymakers when regulatory responses are based in inaccurate assessments.

Of course, there are major pitfalls in treating the processes of medical diagnosis as a false positive and false negative minimizing panacea. This is simply because clinical conditions span a wide spectrum when it comes to diagnostic accuracy, integrity and coherence. At the 'tight' end of this spectrum the objective is to accurately diagnose conditions that constitute clear risks to health and well-being (e.g. to quickly and accurately identify the type of venom injected by a venomous creature of the precise type of cancer present). At the 'loose' end of this spectrum (where the problems lie) there is the tendency to treat more systemic and interrelated conditions using simplistic labels and categories that ignore, or downplay, highly complex problems associated with behavioural links and psychological factors. In such cases, diagnosis can be far more subjective (relying on arbitrary scoring systems subject to significant revision) and can underplay the importance of complex and socially influenced 'driver' factors relative to 'driven' factors (symptoms). Whilst clinicians recognize the limitations of applying highly simplistic diagnostic labels in such circumstances the patient may be expecting a simplistic conclusions and a comparably oversimplified 'silver bullet' solution. Importantly, in the case of the NDIS, accurate diagnosis of mental illness (which is also part of the NDIS) is generally more difficult than physical disabilities and subject to more 'churn' (i.e. people often have periods of mental illness at different points in their life and continuity of care may be difficult to secure through the NDIS).

Consequently, in emphasizing the analytical utility of adopting signal processing methods that seek to estimate the likelihood of obtaining false positive and false negative test results we recommend that lessons be extracted from the relatively unproblematic 'tight' end of the clinical diagnosis spectrum (e.g. cancer detection) and that the ambiguities and pitfalls of diagnostic practice at the 'loose' end of the spectrum be born in mind – especially given the remit of the NDIS (which will inevitably encounter major challenges in diagnostic accuracy and utility in such situations). Indeed, it is preferable to apply these signal processing methods to the evolving administrative procedures used by the NDIS (which can be framed as hypotheses being tested) rather than simply to the medical and health conditions of current and candidate NDIS users.

The signal processing and machine learning methods we consider are derived from Bayesian principles but have the major advantage that, in focusing on test accuracy (as the basis of learning algorithms), they express both the information update process and
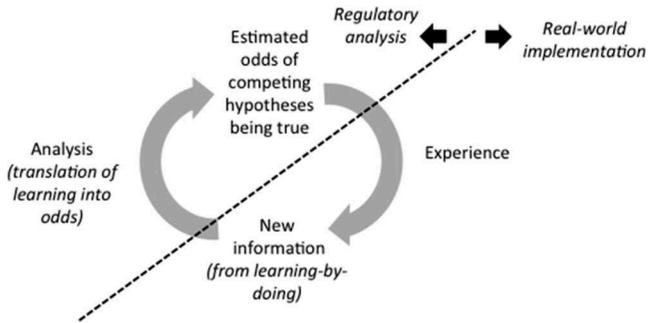
Figure 2: How learning-by-doing in regulation can be expressed in Bayesian terms
Source: Matthews (2015).

the various permutations of test result accuracy in a readily understandable manner. This means that using these methods in a public policy and public management context avoids one of the major impediments to using Bayesian methods in governance: an excessive reliance on bespoke and complex mathematical formulations that are very hard for non-specialists to grasp and difficulties in communicating results in ways that non-specialists can grasp easily (Matthews, 2015).

If regulatory stances can be expressed as competing hypotheses with binary (true/false) answers then it is possible to apply these signal processing and machine learning methods to address regulatory challenges. We return to this issue in more detail after explaining the logic of these signal processing and machine learning methods. At this stage, it is worth noting that experiments in using structured hypotheses testing as a method for decreasing the costs and lead times in public sector evaluation work have been successful (Mathews and White 2013). Evaluation is of course an important component of adaptive regulation so that initial experimental work indicates that there is the potential to broaden the approach to regulatory stances.

Figure 2 expresses the logic of adaptive regulation as a simple Bayesian learning cycle in which the differential odds of competing hypotheses being true are updated by real world experience in policy implementation.

Table 1 contains the basic analytical taxonomy used in signal processing and machine learning, this is sometimes (usefully) referred to as a 'confusion matrix' by engineers because it draws out the ways in which binary test results can be wrong and, in combination, contradictory and, hence, cause confusion. In a machine learning context based on the use of algorithms this confusion paralyses learning and adaptation. In a policy context, the impact on human judgement can be equally paralysing or can lead to decisions being made that arbitrarily ignore this confusion. This can lock interventions into problematic developmental pathways if not corrected at later stages.

As the confusion matrix highlights, we should prefer regulatory stances (if expressed as competing hypotheses with binary answers) that maximize the true positive rate and

Table 1: The 'Confusion Matrix' used in signal processing and machine learning

Test result	Condition present	
	Yes	No
Positive	(a)	(b)
	True positive rate (TP)	False positive rate (FP)
Negative	(c)	(d)
	False negative rate (FN)	True negative rate (TN)

the true negative rate but that also minimize the false positive and the false negative rates. Whenever there are false positive and false negative test results the response of the regulatory framework is *itself* a risk to effective policy delivery (actions may be taken that are unnecessary or actions that should be taken are not taken).

Using a clinical example, Figure 3 contains an illustration of the significance of test result errors (using data from Gigerenzer 2002). For many people this 'natural frequency' based expression of the situation, which clearly communicates relative scale, is far easier to grasp than the standard Bayesian equation.

In presenting the data in this manner it is clear that a positive test result (in this case for colorectal cancer) means that there is only a 4.8 per cent likelihood (known as the *test sensitivity*) that a particular patient actually has the condition. This is simply because
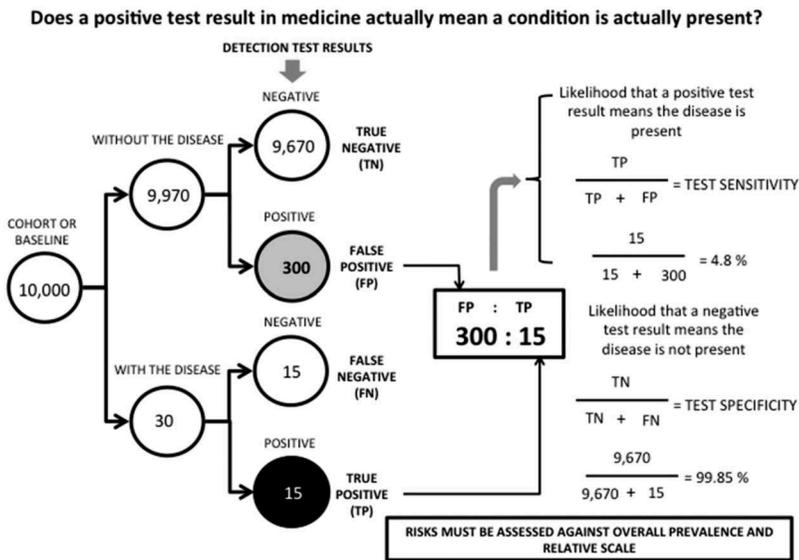

Figure 3: Using natural frequencies to express Bayes rule

Application of a Bayesian hypothesis-testing network to assess the effectiveness of disability services
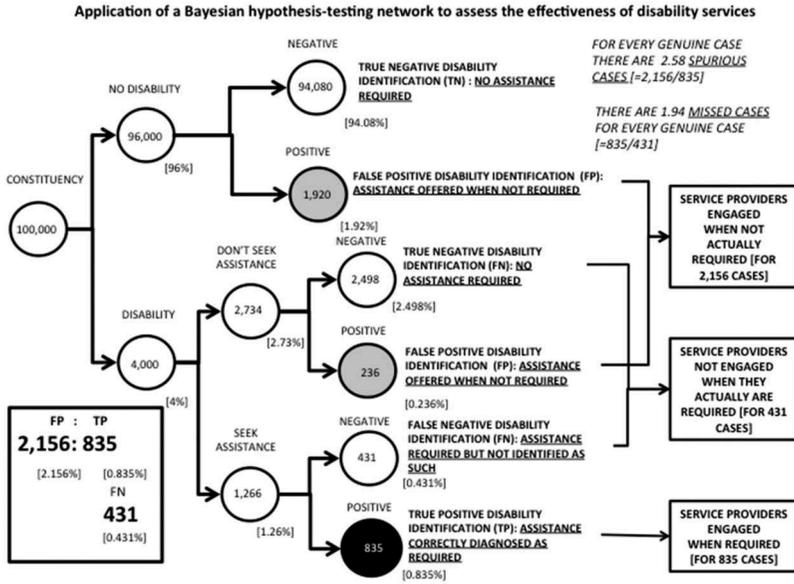

Figure 4: Expressing regulatory test accuracy via natural frequencies

the 3 per cent false positive rate applied to the 9,970 in every 10,000 people who in statistical terms do not have the disease results in 300 cases of false positives relative to 15 true positives. Hence, for an individual patient one must consider the implications of this ratio of 300 false positives against 15 true positives (the odds from which favour a particular test result being a false positive). This highlights the way in which the overall prevalence of a disease in the population, combined with the rates of true and false positives (and true and false negatives) in test results, generates this gap between a naïve interpretation of a particular test result and a more thoughtful and evidence-based interpretation.

Figure 4 contains an illustration of how this natural frequency approach can be used in a regulatory context. This illustration shows how the diagnostic test can detect the inappropriate use of a service as well as individuals who require services but are not receiving them. Note that the particular values given here are purely for illustrative purposes. This illustrates the way in which a very small (1.92 per cent) false positive rate in service applicability can result in a large number of cases treated as being required when they are not required. This diverts scare resources to dealing with cases that should not, in reality, require assistance.

It is easy to see how this use of simplified Bayesian signal processing and machine learning concepts provides a robust and intuitively straightforward basis for assessing aspects of the efficiency and the effectiveness of regulatory frameworks for interventions
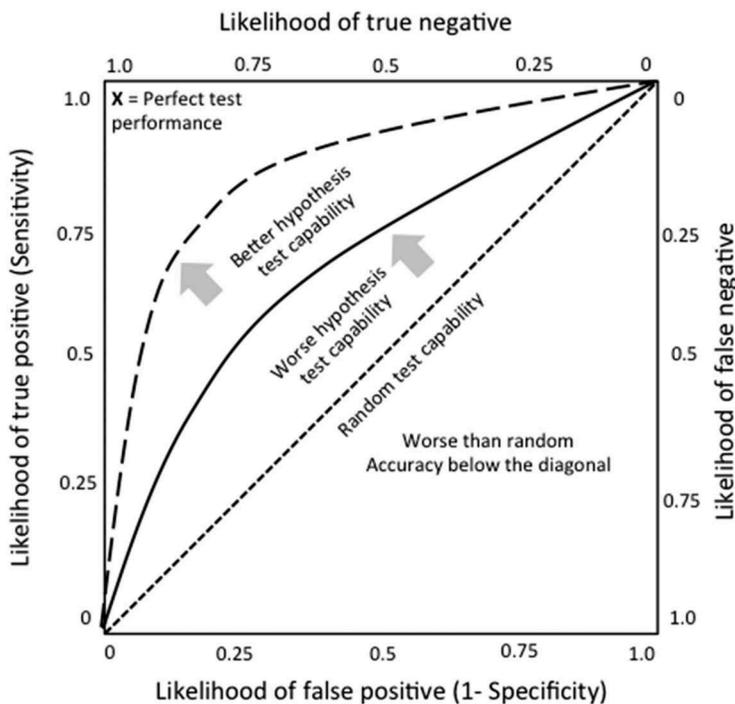
Figure 5: Diagnostic capability curves in a signal processing framework
Source: Figure 7 in Matthews and Kompas (2015).

like the NDIS. The approach makes clear where problems caused by test inaccuracies lie, highlights the implications for regulatory response decisions and provides a basis for measuring historical changes in diagnostic capability in a regulatory framework. The more accurate the diagnostic capability becomes the more effective the regulatory function.

This issue of diagnostic capability is formally expressed in signal processing and machine learning in the following manner (see Figure 5). For historical reasons this is referred to as the *receiver operating characteristic* curve (an ROC curve in short). An ROC curve plots the false positive rate against the true positive rate and was originally developed to assess the abilities of radar operators in World War II. As a diagnostic tool it provides a useful means of measuring the accuracy of test results in a robust and coherent manner. As such, ROC curves reflect the principles behind the use of randomized control trials (RCTs) in public policy – but in a more generally applicable framework (indeed ROC curves are used in medicine to assess the adequacy of RCT results). (For a useful overview of the use of ROC curves in a range of contexts, see Swets, Dawes, and Monahan 2000).
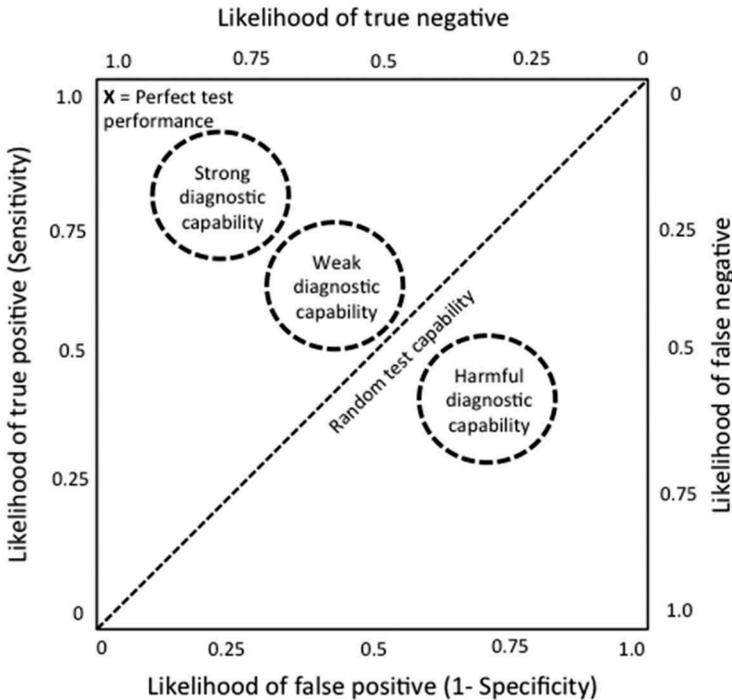
Figure 6: Characterizing regulatory test capabilities using a signal processing framework
Source: Based on Figure 8 in Matthews and Kompas (2015).

The best possible performing hypothesis test lies in the top left-hand corner (a test that is 100% sensitive and has a zero false positive rate). Random test results lie on the diagonal (e.g. someone guessing the toss result of a coin would expect to eventually end up at the 0.5, 05 point in the middle of the diagonal). Test results that are worse than random lie below that diagonal (thus providing a particularly useful diagnostic).

In a public policy context, the potential to waste public funds increases further that capabilities lie from the ideal diagnostic point in the top left-hand corner of the ROC space. Particular test capabilities can be represented as curves in this space: the further above the diagonal and the greater this curvature the more reliable the hypothesis test is. Shifts in capability over time can be reflected as shifts in these curves.

This diagnostic framework can be translated into a governance and regulatory context by characterizing the test accuracy of different approaches. Figure 6 illustrates this principle by identifying three regulatory regimes of differing diagnostic capability: a strong regulatory regime, a weak regulatory regime and a harmful regulatory regime. The latter lies below the diagonal line and hence reflects a situation in which test accuracy is worse than random in the sense that the results achieved are negatively

correlated with the actual results. The framework provided in Figure 6 provides a useful basis for thinking about practical ways of designing and delivering an adaptive regulation framework for the NDIS and similar interventions. This type of framework has the major advantage that it readily maps across to the monitoring and evaluation methods that are used in the public sector. This is because changes over time in the rates of true positive, false positive, true negative and false negative test results provide a key way of measuring cost-effectiveness and demonstrating changes in cost-effectiveness over time (via calculating what it has costed to deliver changes in test accuracy in these signal processing terms). The scope for implementation for large and complex interventions like the NDIS is the focus of the next section of this paper.

EMBEDDING LEARNING AND ADAPTATION INTO GOVERNANCE AND REGULATION ARCHITECTURES

Learning methodologies, such as the Bayesian approach outlined above, are unlikely to be effective if they are not embedded in a broader structure (and culture) of learning and adaptation. By this we mean a management or institutional 'architecture' that encourages learning and adaptation.

Adaptive management fosters learning process in 'helping to inform decision-making, and management contributes to learning by the use of interventions' (Williams and Brown 2014). Whilst a range of frameworks exist for adaptive management, we draw on Williams and Brown (2014) (see also Carey and Harris 2015). Williams and Brown (2014) framework has two principle phases: (a) the deliberative (or planning) phase and (b) an iterative phase, which takes the elements and results of the deliberative phase, folding them into a sequential process of decision-making and learning. This second iterative phase uses elements of the planning phase in an ongoing cycle of learning.

Williams and Brown draw broadly on the seminal work of Argyris and Schon (1978) into organizational learning and experimental learning. Argyris and Schon argued organizations can readily undertake single-loop learning, where a mistake is made and then corrected. Double-loop learning, however, requires organization to 'learn' in response to errors not just by correcting them, but by correcting the norms and cultures that lead to them. Hence, a true learning architecture requires organizations to reshape norms, culture or policies (Argyris and Schon, 1978). However, double-loop learning is difficult because organizational cultures and norms are historically embedded, learned and challenging to shift (Giddens 1984; Linde 2008).

To guide the implementation and delivery of the NDIS, a double-loop adaptive learning cycle could be established between a diagnostic unit (undertaking and fostering learning through a Bayesian monitoring and evaluation framework) and the governance architecture already in place (namely the Department of Social Services and the National Disability Insurance Agency). Figure 7 provides a conceptual diagram of double-loop learning. These two agencies have principal responsibility for 'steering'

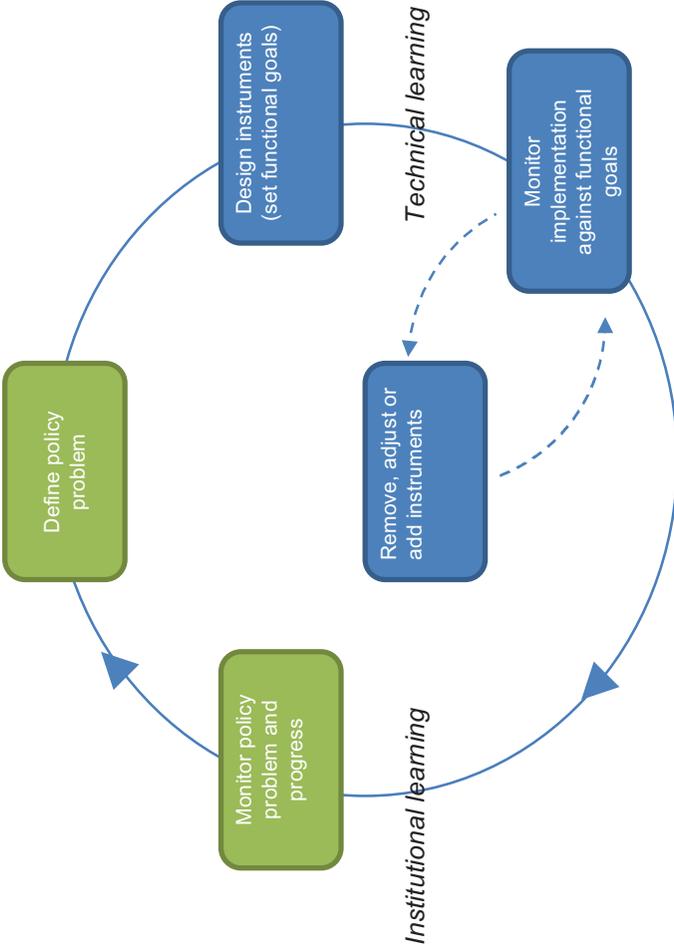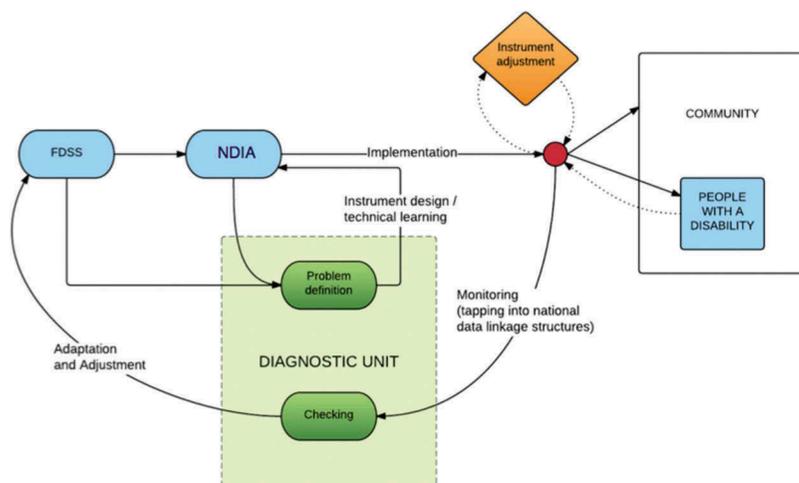**Figure 7: Adaptive management learning cycle (Carey and Harris 2015).**

Figure 8: An adaptive architecture for the NDIS

NDIS markets and service provision. Figure 8 illustrates how a diagnostic unit that utilizes Bayesian analysis could be integrated into a broader adaptive learning govern-ance structure for the NDIS. As noted earlier, signal processing methods could be applied to the evolving administrative procedures and implementation instruments used to guide the NDIS, helping to create the double-loop learning cycle.

Thus, in addition to standard outcome targets, functional targets can be established concerning:

a. The functioning of the diagnostic unit. This would be done by measuring its ability to improve test results and more accurately 'diagnose' those receiving unnecessarily (or wrong) services and those requiring but not receiving services.
b. The instruments used to guide market development and responsiveness to user demand. This would be evident in monitoring data and information from service delivery organizations, peak bodies and other organizations operating in the disability space.
c. Assessing and demonstrating cost-effectiveness using this Bayesian framework – an approach particularly well suited to expressing the advantages of experi-mental and adaptive intervention architectures.

As Argryis and Schon (1978) note, few organizations can do double-loop learning. In the context of the NDIS, with its large numbers of administrators, this challenge should not be underestimated. However, the establishment of functional targets and the proposed diagnostic unit (using signal processing methods) could support double-loop learning by bringing tacit knowledge to the surface, where it can be

explicated and reshaped (i.e. by encouraging reflexive monitoring or reflexive practice) (Giddens 1984). Adaptive management strategies often fail to achieve double-loop learning because this deeper learning can mean individuals must challenge not just norms and values, but their superiors. By establishing these alternate structures, organizational processes (and by virtue, organizational assumptions and norms) are put at the forefront of learning cycles, rather than organizational mission per se.

Hence, while challenging, if the adaptation and learning cycles are functioning appropriately, service delivery will become more effective and efficient over time – despite initial risks associated with experimentation. A Bayesian monitoring and evaluation framework will draw out this aspect of performance – productivity is best measured via identifying changes in rates of wasted effort over time. Hence, an emphasis on test accuracy reveals aspects of wasted expenditure.

Successfully articulating this 'binary' approach to the analysis of competing hypotheses using signal processing and machine learning tools opens up a rich new avenue for people working on the ground at the coalface of service delivery to suggest imaginative new hypotheses based on their accumulating experience and rich tacit knowledge – both diagnostic and as experimental interventions. Indeed, a Bayesian monitoring and evaluation framework directly encourages and facilitates this aspect by highlighting the ways in which these bottom-up contributions can increase the cost-effectiveness of the intervention.

The conducive and vibrant comparative testing environment that we are alluding to encourages this 'bottom'-up approach by providing the concepts, methods and values needed. Indeed, proponents of experimentation in policy have argued for strong systems at the community level that have both local problem-solving capacity and the ability to feed information upwards (both qualitative and quantitative) where required (Dorf and Sabel 1998; Noonan, Sabel, and Simon 2009; Sabel 1995). Noonan, Sabel, and Simon (2009), for example, argue for local panels of volunteers and experts to play a role in implementing and maintaining systems of service provision. Similarly, Braithwaite (Forthcoming) has suggested that rather than generating top-down compliance models (as seen with new public management reforms (Considine 2003)), experimentalism offers a method of reform that encourages authorities to engage with 'the local': 'incremental improvements on the basis of learning, experimenting with innovation, and strengthening professional understanding and practices'.

Without supportive architectures to manage, govern, regulate and guide implementation and ongoing management history shows us that reforms which utilize public service markets tend to become rigid through ever-growing layers of compliance and accountability checks (Considine 2003; Considine, O'Sullivan, and Nguyen 2014; Considine and Lewis 2012). Here, governments become increasingly concerned with regulating risk at the service delivery and client level. This has been a major critique of new public management (Hood and Dixon 2015; Osborne 2010). We suggest that an openness to risk during implementation might be key to moving beyond the limitations

of new public management, which is increasingly seen as ineffective and outdated (Hallsworth 2011; Osborne 2006). We contend that governments need to embrace experimentation, learning and adaptation and we have proposed a practical and conceptually robust means of delivering such innovative approaches. We recognize that this is challenging. System rigidity has emerged in a range of contexts under new public management approaches (Considine and Lewis 2012), indicating that it may be a natural end product of governments managing political risk and accountability. However, we should resist the notion that this is a path-dependent outcome (Kay 2006). If we, as management scholars and practitioners, wish to move beyond the limitations of new public management, we must seek out the new 'ways of doing' (Osborne 2010).

CONCLUSION

The successful implementation of the NDIS confronts two of the largest, and currently unsolved, challenges of public administration – the use of market arrangements in service delivery and the ability to build flexibility and learning into public management structures while still maintaining a level of control. We argue that the creation of a plural and pluralist state (Osborne 2010) requires governments to develop greater openness to risk and policy experimentation during implementation. In this paper we have begun the process of specifying the type of adaptive system architecture that could encourage double-loop learning, thereby supporting adaptation and learning in public service provision in a post-new public management world. While we do not underestimate the difficulty of establishing adaptive learning in practice, without such practices complex social reforms are likely to become rigid and unworkable for system actors over time.

One weakness of the new public management stance is that it has framed cost-effectiveness and productivity in the public sector simplistically via comparison with the private sector (thus driving privatization and quasi-market-based solutions) – rather than articulating a productivity agenda that reflects what is distinctively *different* about what the public sector does. The solution we propose here is directly relevant to capturing the parameters that are distinctive to what the public sector does – for example, managing the uncertainties and risks that the private sector cannot cope with (Matthews and Kompas 2015) – and therefore points the way to ways of overcoming the limitations of the new public management model.

ACKNOWLEDGEMENTS

for their insights and wisdom. Additional thanks to Nic Carey for assisting in the development of the adaptive and learning framework (Figure 8).

NOTE

1 The authors are grateful to Ian Marsh for drawing attention to the importance of developing effective ways of assessing and demonstrating the value-for-money for experimental and adaptive modes of governance.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

REFERENCES

6, P. 1997. *Holistic Government*. London: Demos.

Argyris, C., and D. Schon. 1978. *Organisational Learning: A Theory of Action Perspective*. Canada: Addison-Wesley Publishing.

Australian Productivity Commission. 2011. *Disability Care and Support: Productivity Commission Inquiry Report*. Melbourne, Vic: Productivity Commission.

Baker, A. 2012. *The New Leviathan: A National Disability Insurance Scheme*. Sydney: Centre for Independent Studies.

Barrett, S. M. 2004. "Implementation Studies: Time for a Revival?" *Public Administration* 82: 249–262.

Bonyhady, B. 2014. "The NDIS Vision: Delivering the Plan." St Laurence National Conference, Melbourne, August 11–12.

Braithwaite, J. 2008. *Regulatory Capitalism: How It Works, Ideas for Making It Work Better*. Cheltenham, UK: Elgar Publishing.

Braithwaite, V. Forthcoming. "Removing Punishment from Care and Protection." In *21 Century Public Management: The Experimentalist Alternative*, edited by I. Marsh and C. Sabel. New York: Columbia University Press.

Carey, G., and B. Crammond. 2015a. "What Works in Joined-up Government? an Evidence Synthesis." *International Journal of Public Administration* 38 (13–14): 1020–1029.

Carey, G., and B. Crammond. 2015b. "What Works in Joined-up Government? An Evidence Synthesis." *International Journal of Public Administration* 1–10. doi:10.1080/01900692.2014.982292.

Carey, G., B. Crammond, and T. Riley. 2015. "Top-down Approaches to Joined-up Government: Examining the Unintended Consequences of Weak Implementation." *International Journal of Public Administration* 1–12. doi:10.1080/01900692.2014.903276.

Carey, G., and P. Harris. 2015. "Developing Management Practices to Support Joined-Up Governance." *Australian Journal of Public Administration*. Advance online publication. doi:10.1111/1467-8500.12169.

Carey, G., P. McLoughlin, and B. Crammond. 2015. "Implementing Joined-up Government: Lessons from the Australian Social Inclusion Agenda." *Australian Journal of Public Administration* 74: 176–186. doi:10.1111/1467-8500.12096.

Considine, M. 2003. "Governance and Competition: The Role of Non Profit Organisations in the Delivery of Public Services." *Journal of Political Science* 38: 63–77.

Considine, M., and J. M. Lewis. 2012. "Networks and Interactivity: Ten Years of Street-level Governance in the United Kingdom, the Netherlands and Australia." *Public Management Review* 14: 1–22. doi:10.1080/14719037.2011.589613.

Considine, M., S. O'Sullivan, and P. Nguyen. 2014. "Mission Drift?: The Third Sector and the Pressure to Be Businesslike: Evidence from Job Services Australia." *Third Sector Review* 20: 87.

Crozier, M. 1964. *The Bureaucratic Phenomenon*. Chicago: University of Chicago Press.

Dorf, M., and C. Sabel. 1998. "A Constitution of Democratic Experimentalism." *Columbia Law Review* 98: 267–473. doi:10.2307/1123411.

Giddens, A. 1984. *The Constitution of Society*. Cambridge, UK: Polity.

Gigerenzer, G. 2002. *Calculated Risks: How to Know When Numbers Deceive You*. New York: Simon & Schuster.

Hallsworth, M. 2011. *System Stewardship*. London: Institute for Government.

Hill, M., and P. Hupe. 2009. *Implementing Public Policy*. 2nd ed. London: Sage.

Hood, C., and R. Dixon. 2015. "What We Have to Show for 30 Years of New Public Management: Higher Costs, More Complaints: Commentary." *Governance*. doi:10.1111/gove.12150.

Joint Standing Committee on the National Disability Insurance Scheme. 2014. *Progress Report on the Implementation and Administration of the National Disability Insurance Scheme*. Canberra: Commonwealth Government of Australia.

Kay, A. 2006. *The Dynamics of Public Policy: Theory and Evidence, New Horizons in Public Policy*. Cheltenham, Glos, UK ; Northampton, MA: Edward Elgar.

Kickert, W., and J. Koppenjan. 1997. "Public Management and Network Management: An Overview." In *Managing Complex Networks: Strategies for the Public Sector*, edited by W. Kickert, E.-H. Klijn, and J. Koppenjan, 35–60. London: Sage.

KPMG. 2014. *Interim Report: Review of the Optimal Approach to Transition to the Full NDIS*. Canberra: KPMG.

Larkin, M., and H. Dickinson. 2011. *Personalisation: What Will the Impacts Be for Carers?* Birmingham: TSRC.

LeGrand, J. 2007. *Delivering Public Services Through Choice and Competition: The Other Invisible Hand*. Princeton: Princeton University Press.

LeGrand, J., and W. Bartlett. 1993. *Quasi-markets and Social Policy*. London: Macmillan.

Linde, C. 2008. *Working the past: Working the Past: Narrative and Institutional Memo*. Oxford: Oxford University Press.

Mathews, M., and G. White. 2013. "Faster, Smarter and Cheaper: Hypothesis-Testing in Policy and Program Evaluation." In *European Evaluation Society Newsletter (Evaluation Connections),* 13–14. Prague: European Evaluation Society.

Matthews, M. 2015. "How Better Methods for Coping with Uncertainty and Ambiguity Can Strengthen Government - Civil Society Collaboration." In *Designing and Implementing Public Policy: Cross-sectoral Debates*, edited by G. Carey, K. Landgvot, and J. Barraket. New York: Routledge.

Matthews, M., and T. Kompas. 2015. "Coping with Nasty Surprises: Improving Risk Management in the Public Sector Using Simplified Bayesian Methods." *Asia & the Pacific Policy Studies* 2: 452–466. doi:10.1002/app5.100.

NDIS. 2014a. *National Disability Insurance Scheme Website*. Canberra: Commonwealth Government of Australia.

NDIS. 2014b. *NDIS Evaluation - Information about the Evaluation of the Trial of NDIS*. Canberra: Commonwealth Government of Australia.

NDIS. 2015. *Proposal for a National Disability Insurance Scheme Quality and Safeguarding framework: Consultation Paper*. Canberra: Commonwealth Government of Australia.

Nevile, A. 2013. "The Curse of Accountability: Assessing Relationships in the Delivery of Employment Services." *The Economic and Labour Relations Review* 24: 64–79. doi:10.1177/1035304612474214.

Noonan, K., C. F. Sabel, and W. Simon. 2009. "Legal Accountability in the Service-based Welfare State: Lessons from Child Welfare Reform." *Law & Social Inquiry* 34: 523–568. doi:10.1111/lsi.2009.34.issue-3.

O'Flynn, J. 2013. "Crossing Boundaries: The Fundamental Questions in Public Management." In *Crossing Boundaries in Public Management and Policy: The International Experience*, edited by J. O'Flynn and J. Wanna, 11–44. London: Routledge.

O'Flynn, J., F. Buick, D. Blackman, and J. Halligan. 2011. "You Win Some, You Lose Some: Experiments with Joined-up Government." *International Journal of Public Administration* 34: 244–254. doi:10.1080/01900692.2010.540703.

Osborne, S., ed. 2010. *The New Public Governance*. New York: Routledge.

Osborne, S. P. 2006. "The New Public Governance?" *Public Management Review* 8: 377–387. doi:10.1080/14719030600853022.

Patton, M. 2010. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York, NY: Guilford Press.

Pollitt, C. 2003. "Joined-up Government: A Survey." *Political Studies Review* 1: 34–49. doi:10.1111/psr.2003.1.issue-1.

Rhodes, R. 1997. *Understanding Governance*. Buckingham and Philadelphia: Open University Press.

Sabel, C. 1995. "Learning by Monitoring." In *Handbook of Economic Sociology*, edited by E. Smelser and R. Swedberg, 135–165. Princeton: Princeton University Press.

Swets, J., R. Dawes, and J. Monahan. 2000. "Psychological Science Can Improve Diagnostic Decisions." *Psychological Science in the Public Interest* 1: 1–26. doi:10.1111/1529-1006.001.

Williams, B. K., and E. D. Brown. 2014. "Adaptive Management: From More Talk to Real Action." *Environmental Management* 53: 465–479. doi:10.1007/s00267-013-0205-7.

Williams, I., T. Bovaird, H. Brown, K. Allen, H. Dickinson, J. Kennedy, and J. Glasby. 2012. "Designing Whole-systems Commissioning: Lessons from the English Experience." *Journal of Care Services Management* 6: 83–92. doi:10.1179/1750168713Y.0000000012.

Wong, G., T. Greenhalgh, G. Westhorp, J. Buckingham, and R. Pawson. 2013. "RAMESES Publication Standards: Realist Syntheses." *Journal of Advanced Nursing* 69: 1005–1022. doi:10.1111/jan.12095.

Wong, G., T. Greenhalgh, G. Westhorp, and R. Pawson. 2012. "Realist Methods in Medical Education Research: What are They and What Can They Contribute?: Realist Methods in Medical Education Research." *Medical Education* 46: 89–96. doi:10.1111/j.1365-2923.2011.04045.x.