



07
2016

TOWARDS AN ALTERNATIVE DEVELOPMENT MANAGEMENT PARADIGM?

Cathy Shutt

Towards an Alternative Development Management Paradigm?

Cathy Shutt
University of Sussex

Rapport 2016:07

till

Expertgruppen för biståndsanalys (EBA)

Dr Cathy Shutt is a development practitioner with over 22 years' experience of research and practice within the international development sector. She is particularly interested in how different operations of power help or hinder social change. Cathy is an associate tutor at the University of Sussex. She combines teaching on politics and power in aid institutions with consultancy work focused on monitoring, evaluation and learning for transformation as well as value for money analysis. Cathy has a multi-disciplinary perspective that straddles anthropology, development studies, international relations, political sociology, gender studies, accounting and economics.

This report can be downloaded free of charge at www.eba.se

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

ISBN 978-91-88143-18-1
Printed by Elanders Sverige AB
Stockholm 2016

Cover design by Julia Demchenko

Acknowledgements

This report would not have been possible without financial support from the Expert Group of Aid Studies in Sweden.

I am extremely grateful for Chaz Wyman's help with an earlier draft and Rosalind Eyben's kind encouragement and generous editorial support. Sincere thanks go to Sonja Daltung, Anders Ivarsson, Rebecka Kitzing-Ivarsson, Joakim Molander, Janet Vähämäki and Ray Rist for critical but constructive comments. Special appreciation goes to Kim Forss and Emma Öståker for substantive advice as well as shepherding me through the review process. In addition, I am indebted to Elbereth Donovan, Jonathan Francis, Craig Valters and Pete Vowles who all provided insights that helped me deepen my thinking on several issues. All errors are my own.

Table of contents

Preface	1
Sammanfattning	3
Sammanfattning av resultatet.....	4
Slutsatser	7
Implikationer	8
Summary.....	10
Summary of findings	10
Conclusions.....	14
Implications.....	15
1. Introduction.....	16
1.1 Methodological discussion.....	18
1.2 Outline	20
2. What is Meant by ‘Results Agenda’?.....	22
2.1 Basic concepts and definitions.....	22
2.2 Results based management in international development cooperation	24
2.3 Summary	29
3. Risks Arising from Results Agenda?	30
3.1 Perceptions of risks	30
3.2 Reflections on the justification for practitioner concerns....	33

3.3	Summary.....	37
4.	Alternatives to a Results Based Management Approach	38
4.1	Assumptions about problems and change.....	40
4.2	Assumptions on knowledge, learning and evidence	45
4.3	Assumptions about power, relationships, roles and behaviours.....	49
4.4	Summary.....	51
5.	Emerging Examples of Alternative Approaches	52
5.1	Political economy and power analysis	52
5.2	Theories of social change.....	54
5.3	Monitoring, evaluation and learning.....	57
5.4	Relationships, roles and behaviours.....	64
5.5	Payment by results.....	65
5.6	Summary.....	67
6.	Implications for Organisational and Institutional Change....	69
6.1	Policymaking.....	69
6.2	Monitoring, evaluation, learning and performance management systems	70
6.3	Financing and procurement	72
6.4	Decentralisation and human resource policies.....	73
6.5	Public Communications.....	75
6.6	Summary.....	76
7.	Conclusions and Implications.....	78
7.1	Problematic assumptions.....	78
7.2	‘Alternative’ ideas and tools	79

7.3	Challenges putting ideas into practice	79
7.4	The need for institutional and organisational change	81
7.5	Implications	82
	References	84
	Abbreviations.....	102
	Results and Evaluation Glossary	103
	Annex 1: Politically Smart Locally Led Case Study	
	Summaries	109
	Annex 2: Useful Reading	112
	Previous EBA-reports.....	115

Preface

Measuring the efficiency of public policy in terms of results is a challenging task in general. Development cooperation is no exception in this respect. On the contrary, there are several factors that suggest the opposite, the most important being that development policies are implemented in other countries.

The Swedish Government has for a long time acknowledged the importance of demonstrating results, both for the sake of aid effectiveness and to inform public opinion. The importance of showing results was stated already in the first government bill on international development cooperation in the early 1960s. Since then, there have been several initiatives from the Government to put a stronger focus on results.

There have also been international initiatives to strengthen aid efficiency. In 2005, the importance of results was highlighted in the Paris Declaration on Aid Effectiveness, in which world leaders agreed to take far-reaching actions to reform the ways they deliver and manage aid. According to this agreement, management and implementation of aid should be carried out in a way that focuses on desired results, and information on results should be collected to improve decision-making and accountability.

After the financial crisis, when many donor countries experienced increasing state budget deficits, the quest for results in development cooperation was accentuated. Also in Sweden, a number of new tools have been developed for results based management and the measuring of results. However, criticism from practitioners of what has been called the new ‘results agenda’, has been growing over time. This criticism is based on a concern that the approaches and tools that have evolved, inspired by New Public Management, are not fit for purpose. But why is this the case? And what are the alternatives?

In this EBA-report, Dr Cathy Shutt, at the University of Sussex, scrutinizes the recent critical debate about results based management in development cooperation, including the main arguments and motives behind the criticism. She argues that the debate is not only about obsessive measurement and reporting of numbers for political accountability, but also a matter of assumptions and how we think about development, evidence and learning. In the report, Shutt explores what could be learned from those who do not just criticise results-based management approaches but also propose alternatives. As a former member of the Big Push Forward – a network advocating

for greater awareness of the politics of the results agenda – she has a good understanding of the debates.

The report is primarily based on the UK’s experiences but it also relates to a broader debate on results based management in international development cooperation. Given the Swedish Government’s recent initiatives to identify and implement management practices beyond New Public Management in the public sector, we hope that the analysis and conclusions presented in this report will stimulate a discussion on management practices and reporting. If development cooperation is to be supported over time by the Swedish taxpayers and make a real difference, there can be no doubt that there is a need to measure and demonstrate results, learn from them and improve the practice over time. The question is not if, but how.

The work on this report has been conducted in dialogue with a reference group chaired by Kim Forss, member of the EBA. The analysis and the conclusions expressed in the report are solely those of the author.

Stockholm, August 2016



Lars Heikensten

Sammanfattning

Alla som på något sätt är involverade i internationellt utvecklingssamarbete vill bidra till resultat och göra skillnad för de fattiga eller marginaliserade människor som insatserna riktar sig till. Det är denna drivkraft som ligger bakom resultatfokuseringen i Parisdeklarationen om biståndseffektivitet från 2005. Givarorganisationerna var då angelägna om att samarbeta med sina partnerorganisationer för att få fram data för att mäta effekter och upprätthålla stödet från allmänheten och politikerna. En del givarorganisationer använde sig av resultatbaserad styrning (RBM, Results Based Management) som en strategi för att öka kunskapen om vad som fungerar och vad som ger effekt. Resultatbaserad styrning sågs också som en metod för att möjliggöra bättre beslutsfattande. Trots dessa goda intentioner är det nu allt fler biståndsaktörer som menar att resultatbaserad styrning inte är lämplig för de komplexa globala utmaningar som världen står inför i dag, som klimatförändringarna och flyktingkrisen.

Den här rapporten utgör ett underlag för diskussion om alternativa metoder för planering, uppföljning, utvärdering och kontraktering inom biståndet. Den är skriven för beslutsfattare och aktörer som är medvetna om att det finns problem med resultatbaserade styrmetoder, men som inte har haft tid att undersöka dem på djupet. I rapporten belyses följande frågeställningar:

- Vad kan vi lära oss av dem som inte bara kritiserar resultatbaserade styrningsmetoder utan även föreslår alternativ?
- Hur skulle institutioner och organisationer behöva förändras för att använda sig av dessa alternativa metoder?

Frågorna har besvarats med hjälp av en genomgång av relevant litteratur på området och ett antal intervjuer med nyckelpersoner. Det analyserade materialet omfattar teoretiskt och empiriskt baserade vetenskapliga artiklar, debattinlägg som publicerats av tankesmedjor samt bloggar.

Sammanfattning av resultatet

Resultatbaserad styrning har utvecklats i ljuset av de politiska reaktionerna på 2008 års finanskris och andra förändringar i den globala politiska ekonomin, såsom framväxten av nya givare. Erfarenheterna hittills har varit blandade.

I Storbritannien har man som svar på allmänhetens skeptiska inställning till internationellt bistånd använt sig av kvantitativa resultatmål, vilket har lett till en förbättrad resultatrapportering. Men det har också inneburit mer kontroll och mindre fokus på lärande. Praktiker har upplevt att de satts att styra program utifrån givarländernas modeller och praxis, där fokus har legat på kvantitativa indikatorer och där man har utgått från att förändring är förutsägbar, linjär och kan kontrolleras. Kritiker bland såväl akademiker som praktiker har hävdat att denna ansats gör samarbetsländerna mindre benägna att följa Parisdeklarationens principer om ägarskap och ömsesidig ansvarsutkrävande. Samtidigt har den här typen av resultatbaserad styrning gjort det svårt för både "biståndsleverantörer" och samarbetspartners att genomföra flexibla program där man kan förhålla sig till och lära sig av förändringar i omvärlden på ett sätt som överensstämmer med rättighetsbaserade strategier.

Enligt vissa kritiker, som exempelvis medlemmarna i nätverket "Big Push Forward" (BPF), har den resultatbaserade styrningen inte bara påverkat det direkt synliga vad gäller vilka riktlinjer och procedurer som används för att styra bistånds samarbeten. Den har även en mer diffus påverkan, som exempelvis att biståndsaktörer uppmuntras att fokusera på biståndets kortsiktiga resultat, som de kan kontrollera, i stället för att beakta mänskliga rättigheter. Den här typen av effekter förstärks när aktörerna tvingas följa förfaranden som de kanske inte själva skulle skriva under på. Medlemmar i BPF uppmuntrar praktiker att arbeta med de positiva aspekterna av resultatagendan, och samtidigt vidta åtgärder för att minska dess negativa effekter. Många nya lärdomar om för- och nackdelar med alternativa styrningsmetoder kommer från insatser som genomförs av grupper som arbetar praktiskt med bistånd. Dessa grupper delar inte nödvändigtvis alla av just BPF:s synpunkter. Det rör sig om grupperingar som "Thinking and Working Politically" och "Doing Development Differently", vissa stora internationella icke-statliga organisationer, en personalgrupp inom "Department for International Development" (DFID) samt en grupp som stöds av "International Institute for Democracy and Electoral Assistance" (IDEA).

Problematiska antaganden

De grupper som är kritiska till resultatbaserad styrning har olika argument för sin kritik. Sammantaget visar de dock på ett antal gemensamma svagheter i de antaganden som ligger bakom RBM-verktygen. Det gäller bland annat föreställningarna om att förändring är något tekniskt, linjärt och förutsägbart. Men kritikerna går längre än så och menar att de som är engagerade i utvecklingssamarbete ofta utgår från olika teorier om vad som orsakar problem, om hur problemen ska lösas och vad som leder till förändring. Det är kanske föga förvånande att påpekandet att olika aktörer har olika uppfattning om målen för biståndsinsatserna har väckt frågor kring kunskap, evidens och lärande. Vem behöver lära sig vad? Var och när ska insatser som genomförs i olika politiska och sociala kontexter bedömas och när ska man kommunicera insatsresultaten? Förutom att kritikerna tar avstånd från premissen att givarnas beslutsfattare alltid bör söka efter lösningar som följer bästa praxis, ifrågasätter de även styrmodellens antaganden om mänskligt beteende. De betonar vikten av informella relationer, makt och sociala normer, som de menar har glömts bort i RBM-modellerna som grundas på uppfattningen att människor drivs av individuella intressen och ekonomiska incitament.

”Alternativa” idéer och verktyg

Det råder ingen brist på idéer om hur biståndsmyndigheter skulle kunna förbättra sin hantering av samtidens politiska, komplexa och interrelaterade nationella och globala problem för att uppnå bättre resultat. Det har utvecklats en rad metoder för planering, uppföljning, utvärdering och resultatstyrning, med utgångspunkt i adaptivt systemtänkande, politisk ekonomi och maktanalyser, liksom sociala relationsteorier – en del nyare än andra. Angreppssätt som bygger på att man stegvis utvecklar lösningar på problem (*problem driven iterative approach/PDIA*) grundas på antaganden om att utveckling och förändring till stor del är politiska, oförutsägbara, icke-linjära och beroende av såväl lokalt ledarskap som kvaliteten på relationerna mellan olika aktörer. Delar av detta tankesätt, särskilt när det gäller politik och makt, har också legat till grund för RBM-kritikernas rekommendationer om hur man ska kunna uppmuntra och bygga relationer i utvecklingssamarbetet som kombinerar tillit och förtroende med kritisk dialog. Detta innefattar utveckling och

främjande av strategier där man bedömer prestationer på grundval av resultat på lång sikt (*outcome*) snarare än resultat på kort sikt (*output*). Vidare framhålls vikten av kunskap om den lokala kontexten, inklusive det politiska läget, och behovet av lärande och anpassning under programmets genomförande.

Utmaningar med att omsätta nya idéer i praktiken

Det finns en del som tyder på att biståndsprogram där de ovan nämnda idéerna används kan leda till bättre resultat. Men ansträngningarna för att integrera politisk-ekonomisk analys och förändringsteori i organisationer som arbetar inom ett New Public Management-paradigm har hittills varit nedslående och därtill verkar en stor del av arbetet sakna genusperspektiv.

Det finns en del goda exempel på kreativa strategier när det gäller att utveckla flexibla resultatindikatorer som kan ändras utifrån de programanpassningar som görs för att möta givarnas behov av resultatuppföljning och lärande. Huruvida man kan enas om sådana strategier och få dem att fungera beror dock ofta på individer. Det är svårt att tillämpa den här typen av strategier på insatser i större skala, på grund av de allt mer komplexa och svåröverskådliga maktförhållandena mellan finansärer och de som styr och genomför program. Löpande kontakter mellan olika aktörer kräver tid och ömsesidig tillit, vilket kan vara svårt att åstadkomma när det politiska trycket gör att många traditionella givare behöver framstå som transparenta och kapabla att hantera risker. Noga anpassade modeller som baseras på prestationsbaserad betalning (*payment by results*) kan i vissa fall skapa möjligheter. Preliminära resultat tyder dock på att man i tillämpningen ofta fokuserar på fel resultat. Det är därför inte sannolikt att sådana modeller kan bli den universallösning för ökat lokalt ägarskap och flexibilitet som en del förespråkare för anpassad programplanering hade hoppats på.

Med tanke på de blandade resultat som beskrivs ovan bör man prioritera att påvisa effektiviteten hos politiskt smarta, adaptiva strategier som möjliggör lärande. En del har dock uttryckt farhågor om att de metoder som används för att generera "evidens" skulle kunna undergräva argumenten för förändring om de kopplas till de antaganden som ligger bakom RBM och New Public Management (NPM). Alla tycker att lärande är viktigt och många är positiva till att man tillämpar olika forskningsmetoder. Diskussionerna kring hur

evidens och lärande ser ut och vad de har för syften är dock fortfarande relativt nya. Dessutom fortsätter debatterna om kopplingen mellan uppföljning, utvärdering och resultatstyrning.

Behovet av institutionella och organisatoriska förändringar

Vi står inför betydande utmaningar när det gäller de institutionella förändringar som krävs för att tillämpa ett alternativt styrningsparadigm i syfte att uppnå bättre resultat och respekt för mänskliga rättigheter. Det räcker inte med nya verktyg för uppföljning, utvärdering och kontraktering.

De som kritiserar RBM hävdar att det finns ett flertal problem som inte kan lösas innan man börjar betrakta biståndsmyndigheter som ett sammanhängande system, snarare än byråkratiska maskiner såsom de beskrivs i NPM-teorierna. Det gäller bland annat att omdana systemen för resultatstyrning och utvärdering så att de uppmuntrar lärande; att utbilda beslutsfattare och allmänheten i biståndsfrågor; att stödja ytterligare forskning om vilka effekter personalpolitik har vad gäller självständighet, tillit och socialförändring; samt att revidera förfarandena för kostnadsberäkningar, kontraktering och riskhantering, liksom principen att få valuta för pengarna.

Slutsatser

De som har kritiserat de dominerande strategierna inom styrningen av biståndet har utvecklat värdefulla redskap och lärdomar som visar att det går att komma förbi vissa begränsande antaganden i samband med resultatbaserade styrningsmetoder och att få bättre resultat. Det har bidragit till att ge kritikerna ett visst momentum. Deras erfarenheter illustrerar dock svårigheterna med att försöka institutionalisera styrningsmetoder som stöder politiskt smarta, komplexa program. Det finns därmed en risk att rörelsen för en politiskt smart styrningsmodell med adaptivt lärande kan bli ännu ett välment, men toppstyrt initiativ som erbjuder nya idéer och verktyg utan att ifrågasätta centrala antaganden i det paradig de vill skifta från. Den typen av förändring kräver att man omvärderar de politiska idéer och antaganden som ligger bakom biståndsmyndighetens politik och styrningssystem. Debatterna om prestationsbetalningsmodeller som

inte lever upp till förväntningarna är en möjlig ingångspunkt för diskussion.

Förslag om normförändringar kommer sannolikt att medföra livliga diskussioner bland samarbetspartners och kolleger. Enligt förespråkare för studier av komplexa system är det själva processen, meningsutbytet och konfliktpunkterna som spelar roll.

Förändring är den oförutsebara konsekvensen av social interaktion. Att fokusera och reflektera kring vad som händer i det vardagliga arbetet och att tala om olika sätt att mäta och förbättra resultat är därför lika viktigt som att göra upp stora planer på institutionella reformer.

Implikationer

Beslutsfattare måste inse de traditionella RBM-verktygens begränsningar, då dessa grundas på orealistiska antaganden som döljer de komplexa politiska utmaningar som biståndsaktörer ställs inför i praktiken. Det finns dock inga universallösningar när det gäller att hitta styrningsmetoder som kan bidra till att åstadkomma och mäta rättighetsbaserade sociala förändringar. Beslutsfattare måste förhålla sig kritiskt till de antaganden som ligger bakom sådana strategier som framhållits som lösningar, men som kan ha varit överskattade, exempelvis prestationsbaserade betalningsmodeller.

Modeller med fokus på organisatoriskt lärande och strategier för att hantera komplexitet är utan tvekan svåra att kombinera med den betoning av kortsiktigt mätbara resultat och de uppföljnings- och styrsystem som är typiska för NPM. Det är därför viktigt att beslutsfattare även tar in sådan förvaltningspolitisk forskning och kunskap som finns inom organisationsteorin, särskilt avseende risk, osäkerhet och komplexa system. Denna forskning och kunskap visar på de potentiella fördelarna med att prova resultatstyrningsmetoder som värderar tillit och lärande framför andra tekniker för resultatmätning.

Om ett paradigmskifte inom biståndsstyrningen ska kunna ske måste beslutsfattarna främja strategier för offentlig kommunikation som kan förklara vad som står på spel. Det behövs bland annat forskning för att undersöka möjligheterna att få allmänheten att beakta biståndets resultat och risker i förhållande till interrelaterade globala frågor som migration och klimatförändringar. Sådana diskussioner skulle kunna belysa de potentiella fördelarna med att

hantera dessa utmaningar med solidariska strategier i stället för strategier som baseras på kortsiktiga resultat.

Summary

Everyone involved in international development cooperation wants to make a difference and for their work to have ‘results’ on the lives of poor or marginalised people. It was this motivation that led to the 2005 Paris Declaration’s focus on results. Donor agencies were committed to working with partners to generate data to measure their effectiveness and sustain public and political support. Some adopted results based management (RBM) approaches to improve learning about what works to enhance impact and, as a consequence, enable better management decision-making. Despite these good intentions, a growing number of development practitioners have argued that results based approaches, embedded in new public management (NPM) theory, are unfit for the complex global challenges, such as climate change and the refugee crisis, facing the world today.

This report is a guide to debates about alternative approaches to the planning, monitoring, evaluating and contracting of international development cooperation. It brings together arguments from different groups and is written for policymakers and practitioners who are aware of the issues but have not had time to explore them in depth. The report explores the following questions:

- What can be learned from those not just critiquing results based management approaches but also proposing alternatives?
- How would institutions and organisations need to change to adopt these alternatives?

These questions are answered through a review of the relevant literature and several key informant interviews. The material analysed includes a range of theoretically and empirically based peer-reviewed articles, opinion pieces published by think tanks and blogs.

Summary of findings

Results based management approaches shaped by political responses to the 2008 global financial crash and other changes in the global political economy, such as the rise of new donors, have had mixed results. In the United Kingdom, for example, the adoption of quantitative results targets in response to public scepticism about aid succeeded in improving results reporting. But it also led to controlling

management practice with little direct impact on learning. Practitioners found themselves managing programmes based on best practice models designed in donor countries, with performance management frameworks that focus on quantitative indicators and assume change is predictable, linear and can be controlled. Academic and practitioner critics claimed this reduced commitments to the Paris Declaration principles of country ownership and of mutual accountability to partner governments. At the same time, such frameworks made it difficult for ‘suppliers’ and government partners to implement flexible programmes able to deal with and learn about real world change in ways consistent with rights-based approaches.

According to some critics, such as members of the Big Push Forward (BPF), the power of RBM goes beyond the visible rules and procedures used to manage aid relationships. It has more diffuse and invisible effects. These include encouraging practitioners to think about development in terms of short-term results they can control rather than human rights. Such effects are exacerbated when practitioners are forced to comply with procedures they may not subscribe to. Members of the BPF advocated that practitioners work with the positive aspects of the results agenda, while acknowledging its politics and taking collective action to reduce its negative effects. Many emerging lessons about the pros and cons of alternative management approaches come from such action that is being pursued by a growing number of practitioner groups, although they do not necessarily share all of the BPF’s views. They include: the Thinking and Working Politically and Doing Development Differently communities; some large International NGOs; a group of staff within the Department for International Development (DFID); as well as a community supported by the International Institute for Democracy and Electoral Assistance (IDEA).

Problematic assumptions

The groups critiquing aspects of results based management are driven by varied motivations; however, their arguments expose a number of common weaknesses in the assumptions underpinning RBM tools. These include the ideas that change is technical, linear and predictable. But they go further and argue that those involved in development cooperation relationships often have different theories about the causes of problems, solutions and pathways to change. Unsurprisingly

perhaps, recognition that different actors have different understandings of the aims of development cooperation interventions has led to questions concerning the nature of knowledge, evidence and learning. Who needs to learn what, where and when to achieve, assess and communicate results for interventions implemented in different political and social contexts? As well as eschewing the notion that donor policymakers should be the principal learners seeking best practice solutions, these critics also dispute NPM's assumptions about human behaviour, highlighting the importance of informal relationships, power and social norms. These are forgotten in RBM models that are based on the notion that people are driven by individual interest and financial incentives alone.

'Alternative' ideas and tools

There is no shortage of ideas about how development cooperation agencies could enhance the way they address the political, complex and interrelated national and global issues facing the world today to achieve better results. A range of planning, monitoring, evaluation and performance management methods informed by complex adaptive systems thinking, political economy and power analysis, as well as social theory about relationships, have been developed, some of which are newer than others. Approaches such as problem driven iterative adaptation (PDIA) are based on assumptions that much development and change is political, unpredictable, non-linear and relies on local leadership and the quality of relationships. Some of this thinking, particularly related to politics and power, has also informed RBM critics' recommendations for how to encourage more critically reflexive trust-based development cooperation relationships. These include the development and promotion of approaches that assess performance based on outcomes rather than outputs. The importance of contextually situated political knowledge and the necessity of learning and adapting during programme implementation are emphasised.

Challenges putting ideas into practice

There is some evidence that development programmes informed by the above ideas can have better than expected results. Yet efforts to mainstream the use of political economy analysis and theories of change in organisations operating within a NPM paradigm have been disappointing, and much of the work appears to be quite gender blind.

There are some good examples of creative approaches to developing flexible results indicators that can be changed in line with programme adaptations that meet donor performance monitoring and programme learning needs. However, agreeing these and making them work often depends on individuals. Scaling up such approaches is difficult because of the increasingly complex and messy power relations between those funding and those managing and implementing programmes. The relational work of constant negotiations requires time and mutual trust, which are hard to achieve when political pressure means many traditional donors have to appear transparent and able to manage risk. Carefully adapted payment by results type models may occasionally provide possibilities. But preliminary evidence suggests that concerns about political accountability mean they often focus on the wrong results. Thus they are unlikely to provide the panacea for enhanced local ownership and flexibility that some advocates of adaptive programming had hoped.

In view of the mixed results described above, demonstrating the efficacy of politically smart, adaptive approaches that enable learning is a priority. But some are concerned that the methods used to generate such 'evidence' could undermine arguments for change if they are embedded in the assumptions underpinning RBM and NPM. Everyone thinks learning is important and many are accepting of pluralist research methodologies. However, discussions of assumptions concerning the nature and purposes of evidence and learning for different actors are still quite nascent. Moreover, debates about the relationship between monitoring, evaluation and performance management continue.

The need for institutional and organisational change

When it comes to the institutional change required to support an alternative management paradigm to increase results and the

achievement of human rights, considerable challenges lie ahead. New monitoring, evaluation and contracting tools are insufficient.

Those critiquing RBM argue that aid agencies need to be viewed as contingent systems, rather than bureaucratic machines as is assumed in NPM theory before a number of issues can be addressed. These include: rethinking performance management and evaluation systems so they encourage learning; development education for policymakers and the general public; supporting additional research about the effects of human resource policy on autonomy, trust and social change outcomes; and re-conceptualising cost relationships, contracting and risk management procedures, as well as value for money frameworks.

Conclusions

Critics of dominant approaches to development management have generated useful tools and lessons that show it is possible to overcome some of the limiting assumptions associated with the implementation of RBM and to achieve better than expected results. This has helped them to gain traction. Yet their experiences illustrate the challenges of trying to institutionalise management approaches supportive of politically smart, complex programmes in order to transition to an alternative development management paradigm. Thus there is a risk that the politically smart, adaptive learning and management movement could become yet another well-intentioned, top-down initiative offering new ideas and tools without challenging some of the norms and core assumptions of the paradigm it wants to change. The kind of change required calls for the re-evaluation of the political ideas and assumptions underpinning aid agency policy and management systems. Debates about payment for results based models that are not living up to expectations provide a potential entry point for discussion.

Proposing norm change is likely to involve heated debate amongst partners and colleagues. According to some proponents of complexity thinking, these are the sites of struggle that really matter. Change is the unpredictable consequence of social interaction. Hence focusing and reflecting on what is happening in everyday practice and conversations about means to measure and enhance results is as important as making grand plans for institutional reform.

Implications

Policymakers need to recognise the limitations of traditional RBM tools underpinned by unrealistic assumptions that obscure the complex political challenges as experienced by those involved in development cooperation relationships. However, there are no magic bullets when it comes to finding management approaches that will enable the achievement and measurement of rights based social change. Thus policymakers need to interrogate assumptions underpinning approaches promoted as solutions that may have been oversold, such as payment for results models.

Since complexity informed adaptive learning models may struggle to meet expectations of management systems constrained by NPM's bureaucratic norms, policymakers can play useful roles in supporting policy research and recommendations informed by contingency theory. These highlight the potential benefits of testing performance management methods that value trust and learning over other techniques for the measurement of results.

If a shift in the international development cooperation management paradigm is to be realised, policymakers need to promote public communication strategies that explain what is at stake. This will involve research to explore the potential for encouraging the public to think about development cooperation results and risks in relation to interrelated global issues like migration and climate change. Such discussions could highlight the potential advantages of long-term solidarity over short-term results based approaches to tackle these complex and pressing problems.

1. Introduction

In early March 2016, I met with the Monitoring Evaluation and Learning (MEL) Manager of a multi-donor funded programme who wanted advice about if and how she should try to renegotiate the results indicators and targets in the logical framework being used to rate the programme's performance. She was concerned that its focus on quantitative indicators and short-term results, like the number of people reached by the programme, provided an inadequate basis for assessing the progress of a complex international transparency and accountability programme that included grant making, research and advocacy. We agreed we would probably have to keep the indicator for the number of people 'reached'. Even though it did not really mean anything, we knew DFID needed it for political accountability reporting to taxpayers. When it came to other indicators, she explained that the United States Agency for International Development, USAID, had stipulated that they could only be changed if it was possible to find retrospective baseline data.¹ Although these requirements were slightly different from those of the Swedish International Development Cooperation Agency, Sida, one of the other donors, USAID's conditions still had to be met. We discussed the possibility of her advocating for an alternative approach using flexible targets and indicators for measuring and managing performance. This had been discussed in a meeting arranged by a Doing Development Differently (DDD) community of practice I had attended the previous week. She remarked on how interesting it was that, despite frequent reference to the advantages of more adaptive and flexible approaches to programme monitoring and management advocated by the DDD group, so little fed through to practice. We ended by agreeing that she probably needed to further explore whether the donor staff managing the programme and the external evaluators responsible for reviewing performance were sympathetic to the idea of flexible targets before making final decisions.

This story captures the essence of the present paper, namely the tensions between, on the one hand, the desire and need of those working in international development cooperation agencies to make a difference and, on the other hand, the current results and evidence-

¹ A baseline describes the situation prior to a development intervention, against which progress can be assessed or comparisons made.

based approaches used to assess programme success. Responses to issues as diverse as global climate change, the Zika virus, the Syrian refugee crisis, poor governance, gender inequality and institutional reform are all politically and technically complicated. The multitude of different actors involved in programmes like the one described above tend to have varied understandings of the nature and causes of the problems to be addressed and therefore what approaches and capacities are needed to tackle these. Donor staff and other practitioners involved in such programmes may not always know *or* be able to discover ‘what works’ for recipient organisations who themselves may not be in agreement about what to do. Hence such programmes are likely to involve power struggles and take unpredictable change pathways with uncertain outcomes. Yet these disagreements and the uncertainty that can result are often obscured in a programme’s indicators and frameworks like the one the MEL Manager had to complete. Uncertainty also tends to be hidden in official reports for the public that suggest development cooperation agencies have far more control over development and social change processes than they do in practice.

Over the last ten years a number of communities representing different interests and disciplines have been critiquing mainstream approaches to measuring results and managing performance. They have also been testing and advocating for different approaches to development cooperation management that take into account the unavoidable uncertainty of change processes. Their suggestions consider the implications of different understandings and theories of evidence, learning, development and social change.

This report is a guide to the critiques, debates and recommendations made by those who are advocating different approaches. It is a scan of theoretical contributions and practical initiatives undertaken by policymakers and practitioners who believe that the dominant approaches to development management do not support the kind of learning required to respond to the national and global challenges facing international development actors today.

Two key questions are considered:

- What can be learned from those not just critiquing results based management approaches but also proposing alternatives?
- How would institutions and organisations need to change to adopt these alternatives?

Deliberately broad in scope, the study provides a critical summary of key points from more in-depth work that has been undertaken at different times by the various groups who have been pushing back against the sector's focus on short-term results and best practice solutions. These include: the Thinking and Working Politically (TWP) group;² the Doing Development Differently (DDD) manifesto community;³ some large international NGOs;⁴ a group of staff working on SMART rules within DFID;⁵ a group supported by International IDEA working specifically on democracy programmes; a network that focuses on the potential of complexity science in development management; a gender network; and the Big Push Forward⁶ that I helped to convene. In addition to the contributions of these established communities, I also include some by well-known evaluation, gender and participatory development experts such as Ray Rist, Michael Q. Patton, Tina Wallace and Robert Chambers. These individuals have long been arguing for monitoring and evaluation approaches that support the learning needs of different development cooperation actors.

1.1 Methodological discussion

In view of the limited time for the study it was not possible to undertake any meaningful empirical work. As its purpose was to scan relevant debates and to identify issues arising together with relevant examples, my main method was an appraisal of theoretical and empirical literature. I began with key texts related to the use of RBM in international development that have been written subsequent to the study on results based management by Vähämäki *et al.* (2011). These

² Thinking and Working Politically Community of Practice website: <http://www.dlprog.org/research/thinking-and-working-politically-community-of-practice.php> [3 April 2016].

³ Doing Development Differently website: <http://doingdevelopmentdifferently.com> [3 April 2016].

⁴ For example, see Mercy Corps. (2015). Managing complexity: Adaptive management at Mercy Corps. Mercy Corps, [Online], Available: https://d2zyf8ayvg1369.cloudfront.net/sites/default/files/Adaptive%20management%20paper_external.pdf [4 April 2016].

⁵ Wingfield, T. and Vowles, P. (2014). DFID is changing its approach to better address the underlying causes of poverty and conflict – can it work? Guest Post from two DFID reformers. From poverty to power, [Online], Available: <http://oxfamblogs.org/fp2p/dfid-is-changing-its-approach-to-better-address-the-underlying-causes-of-poverty-and-conflict-can-it-work-guest-post-from-tom-wingfield-and-pete-vowles/> [3 April 2016].

⁶ The Big Push Forward website: <http://bigpushforward.net> [3 April 2016].

included three OECD donor peer reviews, four journal articles on performance management systems and two related reports produced by the UK Independent Commission for Aid Impact.

I then moved on to review key texts and the most recent reports produced by members of the communities of practice advocating alternatives to RBM approaches, drawing on older bodies of literature I have previously engaged with where appropriate. This generated links to other relevant publications listed in various bibliographies that I reviewed alongside literature recommended by the study's reference group. As the issues discussed are of great interest to policymakers and practitioners, I devoted considerable time to engaging with experts' blogs and their readers' comments. To deepen my understanding of some of the issues, I participated in two workshops and conducted five informal key informant interviews.

The core activities described above generated data relating to the concerns of different groups and individuals about RBM tools. It also revealed a range of conceptual ideas and methods that are being advocated to mitigate the potentially negative effects of RBM tools. These were analysed to identify similarities and differences, paying particular attention to the implicit and explicit mention of assumptions underpinning results based management tools and the alternatives proposed. The findings from these analytical exercises are summarised in Tables 1 and 2 below and elaborated in the relevant sections. Several reports produced by the Overseas Development Institute, together with previous reading and personal experience, formed the basis of findings and reflections related to the kind of institutional and organisational change that would be required to implement an alternative management paradigm.

Although much of the material explored focuses on the UK, other examples are discussed, and the issues raised are considered to be of generic relevance to development cooperation policy elsewhere.

A note on evidence and objectivity

Views about what is evidence and what quality it needs to be for different purposes are central to conversations about the strengths and weaknesses of the results agenda (Camfield *et al.*, 2014), so they need to be briefly unpacked here. No one challenges the idea that development cooperation policy and practice need to be informed by evidence (Guijt, 2015). Yet views differ concerning what kinds of

evidence are good enough for communicating results to parliamentarians and the public versus what is appropriate for informing policymakers and managers who take decisions in different political contexts and at various stages in programme cycles (Gonsalvez De Asis and Woolcock, 2015). These differences in opinion relate not only to pragmatic methodological issues, but also to different ideologies, values and epistemological beliefs about the nature of knowledge.

The literature reviewed varied with respect to the positionality of the authors and their views on matters including knowledge and evidence. Although these differences are not analysed in detail, I discuss the possible implications. Moreover, my previous membership of one of the networks advocating alternatives, the Big Push Forward, poses the risk of personal values and beliefs influencing my interpretation and approach to this study. To mitigate this I include critical reflexivity on my positionality, both by making different perspectives transparent to the reader as recommended by Camfield *et al.* (2014), and by noting some of the criticisms of the Big Push Forward at various points in the report.

Other limitations

Articles and reports relevant to the research questions are published on an almost weekly basis, thus the report does not claim to be an exhaustive study of all that has been written on the subject of RBM and possible alternatives. What it tries to do, however, is to provide the reader with an overview of the issues being debated together with examples of emerging practice and to consider the potential implications for policymakers.

1.2 Outline

I begin by exploring the history of results based management (RBM) in development cooperation, highlighting how different agencies and commentators variously understand the term. In addition to introducing readers to the basic concepts and rationales underpinning donors' decisions to adopt RBM, I look at practitioner perceptions of the risks that a narrow RBM approach might pose to achieving development and social change objectives. Following a brief reflection

on examples that validate or challenge perceptions of such risks, the report moves on to explore critiques, ideas and lessons emerging from the practice of a growing number of policymakers and practitioners committed to thinking about and doing development management differently. A section that considers implications for organisational and institutional change precedes the conclusion.

2. What is Meant by ‘Results Agenda’?

Despite common reference to ‘results agenda’, the term is rarely explained in conversations amongst development practitioners. This section begins by exploring basic concepts and definitions related to results, as well as what is known about the efficacy of RBM in donor public sectors. It then moves on to look at the history of the adoption of results based management within international development cooperation and the emergence of current ‘results agenda’.

2.1 Basic concepts and definitions

International development results are defined as “the output, outcome or impact (intended or unintended, positive and/or negative) of a development intervention” (OECD, 2010:33). At first glance then, the term ‘results agenda’ relates to supporting the achievement of the objectives of development interventions. Hence, results agenda might be considered to be unambiguously positive. In practice matters are more complicated. Since some development cooperation agencies, for example DFID, define results in terms of what can be easily measured, like the number of children enrolled in school, rather than the right to education, those who support a rights-based approach often use it more pejoratively. Eyben and Guijt (2015:2) contend that judging education in terms of enrolment or measurable learning outcomes risks crowding out a transformational approach to education as a process of empowerment.

Results based management (RBM) is central to various results agenda. It refers to the approaches used: to plan and negotiate objectives, indicators and results targets; to develop contractual responsibilities and methods for monitoring, measuring and reporting results; and to respond to deviations from targets (Binnedjikt, 2001:4). In addition, it covers the analysis of performance data for learning, decision-making and organisational accountability (*ibid*).

Although RBM ideas have been around since the eighteenth century, its recent spread is associated with the era of new public management (NPM) that began in the 1980s (Eyben, 2013). NPM involves applying ideas from the business sector to the management of government bureaucracy and the delivery of services (Haynes, 2015). Some argue the spread of NPM was a deliberate ideological project to privatise public services, others maintain it was a more evolutionary

process of efforts to make public administration more efficient (*ibid*). Regardless of the cause, a key assumption underpinning the use of NPM in public services is normative – that management and organisational science applied in business will be effective in public service organisations.

New public management lacks a singular definition, but central NPM ideas are informed by principal agent theory (Eyben, 2015). This theory assumes that because individual agents or staff are rational actors motivated by self-interest, principals or policymakers need to design organisational structures and performance rules to create incentives that will align staff interests with policy agenda (Eyben, 2013). Thus NPM decentralises formal bureaucratic power by a principal to give frontline manager agents more discretion, while finding new technologies of control such as those encapsulated in results based management tools, for example indicator targets. These are believed to create incentives for frontline staff that will govern their behaviour. In other words, RBM involves the development of performance measurement and management systems with associated tools that generate, store and analyse information to regulate an agents' behaviour in the interests of the principal (Eyben, 2015:24). Some basic assumptions providing a rationale for measuring results as a means to manage and enhance performance through incentivising and rewarding success are presented in the box below.

Box 1: The Power of Measuring Results

- If you do not measure results, you cannot tell success from failure.
- If you cannot see success, you cannot reward it.
- If you cannot reward success, you are probably rewarding failure.
- If you cannot see success, you cannot learn from it.
- If you cannot recognize failure, you cannot correct it.
- If you can demonstrate results, you can win public support.

Kusek and Rist (2004) adapted from Osborne and Gaebler (1992)

This thinking, that has been central to RBM approaches introduced into the public sectors of many donor and recipient countries, has had mixed results in terms of enabling learning from success and failure to aid decision-making and improve impact. There are examples of performance management reforms implemented in certain contexts,

for example Eastern Europe, having positive outcomes on service delivery (Dan and Pollitt, 2015:11). However, in many instances effects are ambiguous or RBM reforms appear not to have worked (Bastoe, 2006; Pollitt and Dan, 2013). In a review of 27 country's experiences of RBM reforms, Perrin (cited Bastoe, 2006:103-105) found that most had struggled to shift from a focus on activities and outputs to outcomes that were beyond their control. Making the necessary changes in organisational cultures and generating data of the right quality for decision-making had also proved difficult (*ibid*). According to findings from a large systematic review of new public management reforms, understanding why different approaches 'work or not' is challenging given different approaches taken to RBM and the varied contexts in which it is implemented (Pollitt and Dan, 2013). I return to the relevance of such conclusions for development cooperation later in this section.

2.2 Results based management in international development cooperation⁷

Results based management ideas, such as management by objectives, began to influence donor approaches to management during the 1980s (Eyben, 2015), but it was not until the turn of the century that they really began to take hold. The advent of the good governance and aid effectiveness agenda that encouraged partnership approaches to building capacity to assess progress towards the Millennium Development Goals (MDGs) signalled an important shift from a focus on inputs and outputs to one on outcomes and impacts (Kusek and Rist, 2004).

Greater attention to outputs and outcomes promised to transform the relationship between monitoring and evaluation in development cooperation. Thus the shift to RBM was greeted enthusiastically by many in the evaluation community (Kusek and Rist, 2004). Though some were concerned that monitoring data were unlikely to be sufficiently robust to prove the contribution of interventions to change, other evaluators viewed the move as an opportunity for evaluation to play a more useful role in enhancing impact (Nielsen and

⁷ This section explores the history of results based management in international development cooperation. It will only be covered briefly here as it has been well covered elsewhere (e.g. by Vähämäki et al., 2011; Eyben, 2013; Manning and White, 2014).

Ejler, 2008). In theory it allowed evaluation to shift from a single study undertaken at the end of a project to a process that would allow evaluators to interrogate monitoring data and ask whether development cooperation was making a difference during implementation (Rist and Stame, 2006). This was anticipated to enable project managers and staff to learn from output level monitoring data and adapt (Mayne and Rist, 2006; Nielsen and Ejler, 2008). According to Mayne (2007), evaluators working with public sector and non-for-profit organisations had a lead role to play in the development of results based performance management systems to strengthen organisational learning and knowledge management.

The 2005 Paris Declaration attempted to shift relations of power between donor and partner countries to encourage local ownership and mutual accountability for achieving development results. It also indicated the need for donors to support the development of recipient government performance management systems. Partner countries were going to need to develop the capacity to measure change in key indicators related to the Millennium Development Goals that tackled development problems like maternal mortality, malnutrition and access to water. This prompted a series of roundtable meetings on how to pursue the Paris results agenda that began some years before the final Declaration was signed (Stern, 2008). Interestingly, and arguably in keeping with ideas of mutual accountability, RBM principles published after a 2004 meeting in Marrakesh explicitly rejected the articulation of intentions in terms of management *by* results.⁸ They opted for a managing *for* development results (MfDR) version of RBM instead. The Sourcebook produced after the meeting stressed that no penalties would be applied for missed targets and encouraged a flexible approach to analysing reasons for failure to inform adaptation.⁹

In practice the extent to which different donors have been able to support the development of RBM systems that live up to the Paris Declaration principles described above has been influenced by a mix of domestic and international events (Gulrajani, 2015), as well as bureaucratic norms (Vähämäki, 2015:135). Some donor organisations have reported reasonable success in terms of shifting focus. Bastoe

⁸ Management for Development Results. (2004). MfDR principles in action: Sourcebook on emerging good practices, [Online], Available:

<http://www.mfdr.org/Sourcebook/1stEdition/4-MfDRPrinciples.pdf> [4 April 2016].

⁹ *ibid.*

(2006), for example, describes the adoption of RBM by the African Development Bank (ADB) as being instrumental in providing it with a greater results orientation. However, his discussion of the experience of the ADB and other donor organisations, including USAID, identifies various challenges to developing effective RBM organisational systems not dissimilar to those noted by Perrin cited earlier. Similar issues are also reported in a number of recent bilateral peer reviews conducted by bilateral donors.¹⁰

In many donor countries overall approaches to RBM have been affected by the aftermath of the 2008 global financial crash. An era of austerity, shifting geo-politics, public perceptions of donors failing in relation to the MDGs and competition from private philanthropists have all taken their toll on donor priorities (Gulrajani, 2015). In the UK, for example, commitment to a partnership approach appears to have become less important than demonstrating risk management and accountability to domestic taxpayers, and this has driven understandings of the UK's 'results agenda' (ICAI, 2015).

In 2010 the focus on results and value for money in the UK that had begun under a Labour government gathered momentum following the election of a rightwing led coalition government (Shutt, 2015). It was then that the effects of RBM began to be felt by practitioners. Although completing logical frameworks had long been part of the proposal negotiation processes with DFID, suddenly indicators and targets became far more important. This reflected that DFID was one of the three bilateral donors that decided to adopt a standard indicator and agency-wide results framework which included quantitative results targets for performance management (Holzapfel, 2016: 8).

Since DFID's explicit focus on results that are reported to parliamentary committees and taxpayers for accountability is mainly at output level, casual references to *the* 'results agenda' tended to have negative associations amongst development practitioners in the UK. Yet, as was illustrated in earlier discussions about interpretations of results based management around the time of the Paris Declaration, there are other results agenda with ambitions beyond measuring and reporting results for political accountability. According to Owen Barder (2012a), a leading commentator on development cooperation, they include learning what works to improve management decision-making and dealing with complexity through mechanisms, such as payment by results, which are meant to provide partners with

¹⁰ OECD bilateral peer reviews (OECD 2010, 2013, 2014).

flexibility in implementation. Similarly, staff of the Ministry of Foreign Affairs in Sweden, Sida and Uganda civil society organisation partners described multiple results agenda in the recent study of a Swedish development cooperation (Brolin, forthcoming). Whereas ministry staff were predominantly concerned with collecting data that would enable them to demonstrate accountability to parliament, Sida and partner non-governmental organisation (NGO) staff also saw the results agenda providing the opportunity for learning to improve accountability and effectiveness to citizens in donor and recipient countries.

A particular practitioner's interpretation and experience of RBM and the extent to which it integrates monitoring and evaluation functions will depend on how it is interpreted by her or his employer organisation. A donor that adopts an agency-wide framework with standard indicators and targets that it uses to *attribute* its support and measure effectiveness, like DFID, will require an extremely robust and top-down approach to individual project and programme performance planning, monitoring and management. Whereas an agency that uses a looser results framework to guide its work and assess its *contribution* to results achieved jointly in partnership, like Sida, may take a more relaxed and bottom-up approach.

Box 2: Effects of Efforts to Assess Performance against Aggregate Results Targets: DFID Example

DFID's proactive attempts to assess performance against aggregate results targets had significant effects on all stages of the programme management cycle and at each level of the organisation. In 2011 country programmes and other units had to make offers related to the quantity of results they would achieve during a strategy period.¹¹ These were then advanced in individual business cases for each programme. Business cases require staff to go through a standard process that includes the analysis of evidence to support programme designs and the estimation of the quantity and quality of results that might be achieved through alternative approaches. They also have to complete economic appraisals.¹² Once agreed these cases are tendered with details of the quantity of some of the results that suppliers are expected to achieve, such as the number of children vaccinated or enrolled in education. Such targets are not fixed, but are often renegotiated post contract and recorded in logical frameworks that show linear relationships between inputs, outputs, outcomes, impacts and risks and assumptions. These targets, as well as value for money indicators, become the focus of performance monitoring and scoring during DFID's annual review process. Performance of every programme, irrespective of value, is reviewed in each year of its life to assess whether it is on track to deliver the benefits at the costs expected in the business case.¹³ Problem contracts are identified and advised to take corrective action. Successful programmes, on the other hand, are often asked to increase their levels of ambition. All annual review data is consolidated and analysed so that the quality of the portfolio can be assessed to ensure results targets and budgets can be met. Project completion reviews are complemented by evaluations in some cases. Evaluation designs and models vary considerably, depending on the attributes of programmes amongst other factors, with performance monitoring data playing varied roles in evaluation processes.¹⁴

¹¹ This process followed DFID's 2010 bilateral and multilateral review.

¹² Examples can be found on the DFID Development Tracker <https://devtracker.dfid.gov.uk>

¹³ DFID. (2015). Department for International Development annual report and accounts 2014–2015. UK: DFID, [Online], Available:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/445698/Annual-Report-2014-2015.pdf [31 March 2016].

¹⁴ This is based on personal experience gained through involvement in a number of different DFID funded programmes.

Shortly after the implementation of the results approach described in Box 2, practitioners from the UK began to mobilise and discuss its possible effects. Even though other donors may not have adopted such stringent results frameworks with quite as severe consequences, interest in an initial Big Push Back against this quite narrow interpretation of RBM, that was commonly referred to as '*the* results agenda', suggested it was not only a British phenomenon (Eyben, 2015). Practitioners and policymakers working for donors and international agencies in many Development Assistance Committee (DAC) donor countries were keen to join the debate. They included practitioners from Sweden. Results measurement might not have been pursued quite as purposefully in Sweden as in the UK, yet Sjöstedt, (2013), Statskontoret, (2013) and Vähämäki (2015) made similar claims concerning RBM tools undermining the strategic direction of Swedish Development Cooperation and its principles of local ownership.

2.3 Summary

On the one hand, it is possible to link results agenda articulated by development cooperation agencies to the Paris Declaration's aims of adopting RBM to help donors and recipients focus monitoring and evaluation lenses on outcomes and impacts. However, on the other hand, understandings and experiences of RBM and results agenda in international development cooperation have been influenced by broader factors. These include global as well as domestic political and economic events that have combined with bureaucratic norms, as well as its new public management theoretical underpinnings. In the case of the UK these factors led to quite narrowly defined results agenda that focused on short-term results. Subsequent sections unpack if and how various communities of practice and individual practitioners went on to interpret and respond to such developments.

3. Risks Arising from Results Agenda?

Regardless of the rationales for different results agenda, in the early part of the decade practitioners expressed concerns relating to the risks of RBM leading to ‘obsessive measurement disorder’ (Natsios 2010:4). Using Ebrahim’s accountability terminology, they were worried that results based systems would be driven by donors’ domestic ‘functional’ accountability to taxpayers at the expense of ‘strategic’ accountability needs such as learning, adapting and being accountable to partners. Sentiments described in the MfDR principles and evaluation debates that had emerged in the wake of the Paris Declaration appeared likely to be undermined. The remainder of this section summarises practitioner perceptions of such risks, the issues they raise and then briefly reflects on whether they appear to have any substantive grounds. The discussion is based mostly on recent experience and examples from the UK.

3.1 Perceptions of risks

Some practitioners articulated their worries about the new emphasis on measuring quantitative results in quite technical terms, for example relating to the methodological challenges associated with measuring increases in democracy; others described them in more political terms. Members of the Big Push Forward viewed results described in terms of quick, tangible outputs as contradictory to rights-based approaches that seek to transform power relations (Eyben and Guijt 2015:9). They argued that the narrowly defined results agenda articulated by DFID was seen to privilege the ‘what’ at the expense of the ‘how’. It implied that development is transactional. Statistics about the costs of donors purchasing malaria bed nets to reduce disease risk suggest that development results relate to donors delivering materials, rather than supporting long-term, potentially transformational change, such as enabling governments to develop long-term malaria reduction strategies.

Additionally, the power of RBM was not just perceived in terms of the visible power of rules and procedures such as the logical framework that is used to generate and communicate indicator information. Neither was it restricted to the use of such information for performance management and steering. Practitioner unease also related to invisible power that operates through ideas underpinning

RBM tools that can have perverse consequences (Eyben, 2013:8). They can be used in ways that influence whose ideas count, while suggesting that development agencies are omnipotent and able to exercise a degree of control over outcomes. A more extensive list of the risks posed by results agenda as perceived by practitioners in 2012 and their causes are listed in Table 1 below.

Table 1 Perceptions of the risks of the results agenda

Risks	Causes of risks
Less strategic, equitable, (Barder, 2012b ¹⁵) and transformational development cooperation (Eyben, 2010)	<ul style="list-style-type: none"> - RBM targets skew priorities by creating perverse incentives to focus on the what: results, rather than the how: rights (Eyben and Gujit, 2015) - The need to spend and achieve results quickly reduces incentives to focus on the poorest and most marginalised people who are more expensive to reach and empower (Barder, 2012b) - Projects are the main unit of analysis rather than portfolios (ICAI, 2014)
Reduced funds for innovative, risky, political or complex programmes (Natsios, 2010; Barder, 2012b; Power and Coleman, 2011)	<ul style="list-style-type: none"> - Technical RBM planning tools that ignore different understandings of problems and uncertainty about change pathways are ineffective for planning, observing and measuring results of innovative programmes (Ramalingham, 2013) - RBM incentivises safe, blueprint programming. There are disincentives to propose risky programmes for fear of penalties following real or imagined (unprovable) ‘failure’
Setting up agency wide results frameworks would not be feasible (Barder, 2012b)	<ul style="list-style-type: none"> - Monitoring data does not involve the counterfactual thinking needed for learning about the extent to which an intervention caused a change (Barder, 2012b). - Incentives to cheat, mean that monitoring is performative and unreliable (Barder, 2012b; Shutt, 2011; Welle, 2013) - Aggregation is difficult (Vähämäki <i>et al.</i>, 2011)

¹⁵ References to Barder (2012b) come from a blog in which he summarised concerns about results agenda. They do not necessarily reflect his personal views. Barder, O. (2012b). Seven worries about focusing on results, and how to manage them. Owen abroad, [Online], Available: <http://www.owen.org/blog/5483> [4 April 2016].

<p>Reinforces a notion of transactional rather than transformational development in minds of the public (Roche, 2012)</p>	<ul style="list-style-type: none"> - The main indicators tracked and reported in public accounts communicate that development is a linear and predictable process that donors can buy, control and deliver to less fortunate people (Roche, 2012) - There is no recognition of the role that rich countries have played in causing problems such as climate change and migration
<p>Reduces the quality of learning for strategic accountability (Ebrahim, 2003; Guijt, 2015) and may result in doing the wrong things (Hughes, 2012)</p>	<ul style="list-style-type: none"> - RBM systems prioritise a decontextualised, technical approach to learning, driven by the desire to identify ‘best practice’ and replicate (Chambers, 2010; Eyben, 2013). Donor desire to learn about ‘what works’ is prioritised over contextualised learning needs (Woolcock, 2013); - The methodologies exclude the voices and knowledge practices of partners and poor people, particularly women (Chambers, 2010; Wallace and Porter, 2013) - RBM creates perverse incentives that contribute to biased evidence and interest in justifying rather than identifying, learning from and reporting failure (Camfield <i>et al.</i>, 2014; Morton, 2009; Picciotto, 2016; TWP, 2014) - RBM focuses on short-term indicators, while ignoring unintended outcomes (Booth and Unsworth, 2014; Bamberger <i>et al.</i> 2016; Jabeen, 2016; Vallejo and Wehn, 2016) - More time spent on regressive learning of how to use RBM tools than on learning for social change (Shutt, 2006)
<p>Disempowered partners</p>	<ul style="list-style-type: none"> - RBM performance management tools such as proposals and logical frameworks reduce ownership, disempower and undermine trust (Wallace <i>et al.</i>, 2006; Win, 2004; Abu Alghaib, 2015) - Technical tools for negotiating and reporting that privilege certain kinds of knowledge are unjust and become a means of transmitting neoliberal management approaches (Ebrahim, 2005; Eyben, 2013; Guijt, 2015; Townsend <i>et al.</i>, 2002; Wallace and Porter, 2013)

High transaction costs (Barder, 2012b)	- Processes for agreeing RBM performance targets, establishing processes for monitoring them and taking corrective steering decisions, as well as aggregating data are time consuming and expensive (Barder, 2012b)
Increasing donor control undermines ownership (Barder, 2012b; Leroy 2012), and other Paris Declaration principles of alignment, partnership and accountability to local actors (Roche, 2015)	<ul style="list-style-type: none"> - Donor targets drive procurement, contracting, indicator choices, performance monitoring and management as well as evaluation. This creates incentives for donors to engage in direct delivery to control results rather than capacity building of local systems (Barder, 2012b) - Incentives to work collaboratively suffer because of increased difficulty of ‘attributing’ results in collaborative work (Vähämäki <i>et al.</i>, 2011) - Over specified inputs reduce the opportunity to adapt as a result of learning and undermine relationships of trust (Booth and Unsworth, 2014)

3.2 Reflections on the justification for practitioner concerns

This section reflects on whether, with the benefit of hindsight, concerns listed in the left hand side column of Table 1 appear justified.

First, DFID has outlined some significant changes in its new aid strategy, such as a greater focus on global issues and national interests, which are political and unlikely to have been influenced by the results agenda.¹⁶ However, an important change related to an increase in transparency and emphasis on results is the decrease in budget support and partnership with recipient governments.¹⁷ Whether less funding going directly to partner governments necessarily reduces local ownership is open to debate. Members of the Thinking and Working Politically (TWP) community of practice involved in the critique of RBM approaches used by DFID have been pursuing local ownership through alternative aid modalities such as ‘arms length aid’ (Booth and Unsworth, 2014). Like civil society programming, this modality is

¹⁶ DFID and The UK Treasury. (2016). UK Aid: Tackling global challenges in the national interest, [Online], Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/478834/O_DA_strategy_final_web_0905.pdf [31 March 2016].

¹⁷ The recent national aid strategy includes intentions to stop budget support so aid can be more targeted. It seems reasonable to link this to the results agenda.

seen as a means to provide local reformers in different sectors with access to resources to take collective action, but without such direct and visible links to donors or reliance on external financing (Pinnington, 2014).

Second, innovation remains an important theme for UK Aid, and a focus on results has not diminished this. Nevertheless, DFID has been seeking alternative funding modalities for innovative programmes. Hence it is possible that politically smart staff are trying to reduce the risks of a focus on results reducing funds available for innovation.¹⁸

Third, when it comes to the feasibility of developing systems to serve various results agenda it appears DFID has made more progress in some areas than others. For example, it regularly collates data from performance monitoring systems to manage portfolio performance.¹⁹ Initially, I found no examples of such data informing any programmatic decisions, which would in any case be rare in public sector organisations (Hummelbrunner and Jones, 2013; Furubo *et al.*, 2013; Vähämäki *et al.*, 2011). But this does not mean such performance data have no role in informing organisational or programmatic decisions. Several studies reviewed during the course of this study, for example Honig (2014), suggest that the academic community are using donor performance data for large-scale empirical studies on factors that influence development cooperation performance.²⁰ Furthermore, a macro, theory-based evaluation of DFID's Empowerment and Accountability portfolio, which is attempting to explore which programmes work, and how and why they work in different contexts, is using data from performance-related annual review processes as a key source.²¹

¹⁸ Key informant interview with a senior DFID policy maker during a recent evaluation.

¹⁹ Pete Vowles Head of Programme DFID's Delivery, "Internally, we look at this kind of information on a monthly basis, with management information presented to our Executive Management Committee analyzing the portfolio in a similar way. These MI reports show the distribution scores for Annual Reviews and Project Completion Reviews and then use the Portfolio Quality Index (PQI) to examine performance across the organisation." Comment on Aid Leap. (2016). I analysed 600 of DFID's annual review and here's what I found, [Online], Available: <http://aidleap.org/2016/03/14/i-analysed-600-of-dfid-annual-reviews-heres-what-i-found/> [20 March 2016].

²⁰ Honig (2014) used a large data set from 9 donors to explore the relationship between donor agency and staff discretion on organisational performance. This report was picked up by some CGDEV experts and used as evidence to promote more discretion. However, contradictions between the way DFID was scored in this analysis compared with 2 recent reviews conducted by the Independent Commission for Aid Impact and the OECD Peer Review process raise questions about the appropriateness of some of the codes used.

²¹ Itad. (2015). Empowerment and Accountability Macro Evaluation. Itad, [Online], Available: <http://www.itad.com/knowledge-and-resources/dfids-macro-evaluations/empowerment-and-accountability-macro-evaluation/> [4 April 2016].

Fourth, while I was unable to locate research on the effect of results data on public attitudes to aid, anecdotal evidence from the UK and the US suggests that occasional, well-evidenced examples of results help senior civil servants sustain political support.²² But such evidence appears to come from single study evaluations, which may not reflect the development of the kind of integrated results based management systems envisaged by Kusek and Rist (2004). For instance, Oxfam, a large INGO that has adopted an agency wide results framework that includes standard indicator data from all projects together with a random sample of impact studies, claims the system has helped to maintain public support (Hutchings, 2014). Yet it has not contributed to learning, which is enabled by parallel systems instead (*ibid*). This is discussed further with regard to the feasibility of single RBM systems serving various learning and accountability agenda in Section 6.

Fifth, related to the point above, DFID's focus on measuring results was not pursued as an alternative to learning. Far from it, from 2010 to 2015 DFID invested hugely in developing evidence and evaluation strategies (OECD, 2014:20). Yet this was not through the development of the kind of RBM systems imagined by evaluators like Rist and Stame (2006). In 2014 an OECD peer review found DFID was commissioning too many non-strategic evaluations and producing numerous research studies that no one had time to read or use (OECD, 2014:20). Furthermore, despite lots of learning and sharing taking place amongst implementers, it is not necessarily supported by results data. A mid-term review of a governance programme in Africa recently noted that despite huge amounts of time being invested in collecting data on the number of people attending government meetings, it was not part of the evidence used in learning and outcome mapping exercises exploring changes in behaviour and relationships. Staff who are part of a 'lean' management unit said they had no time to analyse the quantitative data trends.²³

²² Presentation by DFID's Chief Scientific Officer, Watts, C. (2016). What role for evidence in the new aid strategy? Institute of Development Studies, [Online], Available: http://www.ids.ac.uk/events/research-and-knowledge-at-dfid-what-role-for-evidence-in-the-new-aid-strategy?utm_content=bufferd49f7&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer [31 March 2016]; Former USAID Administrators Shah, R. and Gerson, M. (2016). What do we mean by results based development? Centre for Global Development, [Online], Available: <http://www.cgdev.org/blog/what-do-we-mean-results-based-development-podcast-raj-shah-and-michael-gerson> [31 March 2016].

²³ Personal experience.

Finally, DFID may score well in terms of having established an impressive performance management system that can aggregate very basic performance measurement data, but it has not done so without financial and relational costs (OECD, 2014:20). Several recent reviews reported burdensome procedures with damaging effects on staff relationships, motivation and performance which can be traced back to intense ministerial oversight (ICAI, 2015; OECD, 2014:20; Whitty, 2015). Staff operating at country level had little decision-making power and standard indicators had a negative effect on contextually relevant programming, leading to compliance rather than risk taking. Staff were allegedly so disempowered that they were disinclined to challenge the results culture.²⁴ Moreover, the over specification of inputs and results in many programmes was perceived to jeopardise local ownership, however defined. This prompted some of the critique and testing of approaches to enable results reporting without compromising flexibility and locally led planning discussed in Section 4.

Before concluding this section, it worth stating that overall experience of the results agenda not only relates to the specific approach taken by an organisation, but also to an individual's role, positionality, attitude to and use of tools. Thus despite inevitable challenges mentioned above, many practitioners have found the results agenda extremely helpful. A survey conducted by the Big Push Forward in 2013 revealed that non-governmental organisation (NGO) monitoring, evaluation and learning (MEL) staff found the focus on results raised awareness of the importance of monitoring evaluation and learning within their respective organisations (Whitty, 2015). Similarly, many practitioners like myself deepened our understandings of different and more robust evaluation approaches by engaging in results and evidence debates. Moreover, participants at the BPF Politics of Evidence Conference held in 2013 and Global Partners Governance (2014) are amongst those who claim there is nothing inherent about RBM tools that prevent them being used in ways to enhance social change. However, as the discussion above reveals, political interests, particular ideologies and ideas driving

²⁴ Based on Maxwell, S. (2014). Is teacher pleased? The DAC peer review of UK 2014, [Online], Available: <http://www.simonmaxwell.eu/blog/is-teacher-pleased-the-dac-peer-review-of-the-uk-2014.html> [3 April 2016]; similar sentiments were expressed in Big Push Forward spaces.

institutionalisation of results and performance systems can make such use challenging.

3.3 Summary

This section started with a summary of practitioner perceptions of a broad range of risks pertaining to certain interpretations of results agenda. A review of examples concerning the extent to which they materialised in the UK shows variation across various themes. A focus on results appears to have had some positive effects on organisational monitoring and evaluation capacity that is believed to have helped maintain public support and contributed to certain types of learning for practitioners. However, as of 2014, the focus on results appeared to have increased donor control and transaction costs while influencing relationships and the potential for partner learning in somewhat unhelpful ways. These issues are the focus of the next section which moves on to explore alternative approaches and tools being advocated by critics of the RBM approach, and tested in DFID and other agencies.

4. Alternatives to a Results Based Management Approach

During the last six years, the various communities of practice introduced earlier have advocated concepts and approaches that seek to overcome the most detrimental effects of RBM described in Table 1. They do not take issue with *all interpretations* of results based management; however, they do draw attention to the weaknesses of assumptions underpinning some key RBM tools. These assumptions not only encourage the idea that development is mechanistic and can be managed and controlled, they also drive the belief that it is possible to identify and implement best practice solutions (Booth, 2012b; Carothers and De Gramont, 2013; Eyben, 2015; Ramalingham *et al.*, 2014).

A number of practitioners have turned to complexity science and adaptive systems to help conceptual thinking related to planning, monitoring, evaluation and learning for ‘best fit’ rather than best practice solutions. There is no singular definition of complexity science as it encompasses a broad range of theoretical approaches. But Ramalingham (2013), Chambers (2010), Eyben (2005), Hummelbrunner and Jones (2013), Patton (2011) and Root *et al.* (2015) have all employed key concepts to illuminate weaknesses in the fundamental assumptions underpinning the RBM paradigm. I begin the section with a table that summarises their analysis of the ‘problematic’ assumptions associated with the new public management RBM paradigm, contrasting them with alternatives that have informed the approaches they advocate as more effective means for programmes seeking long-term, locally led results. The discussion then moves on to elaborate on a few examples that demonstrate the application of some of the assumptions from the ‘alternative’ paradigm.

Table 2: Assumptions underpinning approaches to development management

Assumption area	More like the established paradigm	More like an alternative paradigm
Problems	<ul style="list-style-type: none"> - Problems can be identified, are bounded and mutually understood; best practice solutions can be mutually agreed 	<ul style="list-style-type: none"> - Different actors have different understandings of problems and solutions
Change	<ul style="list-style-type: none"> - Linear, proportional, predictable and controllable 	<ul style="list-style-type: none"> - Unpredictable and the result of multiple human interactions and feedback, shaped by politics and power
Knowledge, learning and evidence	<ul style="list-style-type: none"> - It is possible to generate objective evidence and use it to inform optimal policy options and programme plans - Evaluation is driven by learning questions to prove attribution and validate policy options - Rational, behavioural approach to learning that is a response to top-down rules and incentives 	<ul style="list-style-type: none"> - No knowledge is value-free, thus policy decisions are based on partial information and political pressure - Planning is based on consideration of different scenarios in light of understanding of political context that includes participatory analysis and consideration of how history happens - Local learning from participatory monitoring of results, or lack of them is key to real-time learning and adaptation - Evaluation is able to explore fundamental assumptions about social change and unexpected outcomes - Learners use deductive and inductive reasoning. They learn and adapt through behavioural, cognitive and social means

Power, relationships and capacities	<ul style="list-style-type: none"> - Formal between atomised individuals, managed by contracts and rules - Informal relationships and individuals' political savvy and relational skills are unimportant - Capacities are easy to organise to achieve common goals 	<ul style="list-style-type: none"> - Informal relationships, trust and flexibility are important; political and relational skills count - Capacities are distributed so collective action is a challenge - Power is everywhere and relationships are messy - Structured relationships maintain informal institutions such as cultural norms that create inequity as well as challenge them
Roles and behaviours	<ul style="list-style-type: none"> - Managing and controlling to satisfy upward accountability and achieve results - Driven by concerns about efficiency 	<ul style="list-style-type: none"> - Facilitative and trusting, allow discretion and encourage learning and quality assurance - More concerned about effectiveness than efficiency

Adapted from ideas in Chambers (2010: 46–47), Ramalingham (2013) and Root *et al.* (2015)

4.1 Assumptions about problems and change

The alternative ideas and assumptions described in the right hand column of Table 2 now feature centrally in the approaches and tools advocated by the Doing Development Differently and Thinking and Working Politically communities as well as those of some evaluators and participatory practitioners interested in complexity science. Such tools are often framed as overcoming problems commonly associated with the logical framework or log frame. The logical framework is a results-based management tool described as improving the design of interventions. It involves identifying strategic elements (inputs, outputs, outcomes, impacts) and their causal relationships, indicators of change, and the assumptions or risks that may influence success and failure.²⁵

The log frame is frequently criticised for suggesting: that development problems as complex as climate change can be bounded

²⁵ OECD (2010).

and commonly understood; that different actors can get together, agree on and design solutions that bring together necessary capacities to deliver; that these will trigger linear, predictable and controllable change; that individuals involved in relationships learn from and respond rationally to top-down rules; and that relationships and feedback are unimportant in the change process (Mowles *et al.*, 2008; Ramalingham, 2013; Chambers, 2010; Global Partners Governance, 2014; Hummelbrunner and Jones, 2013; Maclay, 2015; Eyben, 2015).

Some development interventions are obviously 'simpler' than others in terms of actors' perceptions of the problems that need addressing, their ability to get together, and what is known about the causal mechanisms that determine whether proposed 'solutions' will work or not in a given context. However, apparently simple interventions are frequently complicated by the nature of relationships amongst those involved, as well as by local political contexts. Political Islam's framing of donor funded vaccination programmes in Nigeria as efforts to sterilise Muslims is a pertinent example of the latter effect (Kleinfeld, 2015:17). Although the causal relationship between the vaccine (input) and reduced incidence of disease (outcome) is well established, it could not be assumed to work in a political environment where the legitimacy of the input was challenged to the extent that populations refused to accept it.

According to Hummelbrunner and Jones (2013), complex adaptive systems theory can help to assess levels of uncertainty, the distribution of capacities needed to approach a solution and the selection of the right management approach. RBM does not inherently prevent such action. Nevertheless its implementation in rigid frameworks with performance indicator targets set too high, for example impact level or too low, for example activity level has undermined its potential to do so (*ibid*). A more fundamental problem is RBM's assumption that it is easy to negotiate shared goals and definitions of results that can be formulated into neat quantitative targets. In practice the need to begin collaborative work by agreeing quantitative targets would likely prove a disincentive to building the kind of relationships that are required to achieve results where there is high uncertainty and divergent perspectives on the nature of a problem and how to tackle it. The 2014 Ebola epidemic is a good example as understanding cultural norms proved crucial to the success of medical interventions, though this was not understood at the outset (DuBois *et al.*, 2015)

Ramalingham *et al.* (2014), Hummelbrunner and Jones (2013) and others involved in the various communities of practice challenging a traditional interpretation of RBM have adapted a framework developed by Dave Snowden to describe the emergent and uncertain evolutionary nature of complex systems.²⁶ Drawing on various disciplines, the framework explores relationships between people involved in interventions, experience and contexts. It is a useful tool for decision-making in complicated social contexts, such as development cooperation relationships.

The framework below is one example of an application of Snowdon's tool that has been developed to help practitioners decide what kind of planning, monitoring, evaluating and learning approaches should be used. Criteria for decisions have been developed that relate to what is known about causal relationships between inputs, outputs and outcomes and understanding of the context.²⁷

²⁶ Snowden, D. (2010). Cynefin framework, [Online], Available: <https://www.youtube.com/watch?v=N7oz366X0-8> [3 April 2016].

²⁷ This figure is adapted from proceedings of a workshop organised by USAID's Innovation Lab, the Institute for Development Studies, MStar and FHI 360, and held at Nesta London in October 2015 (Learning to Adapt: Exploring Knowledge, Information and Data for Adaptive Programmes and Policies, 2015) as well as subsequent work by DFID and ODI on guidance for adaptive programming.

Figure 1. Learning to Adapt: Exploring Knowledge, Information and Data for Adaptive Programmes and Policies

<p>Confidence in Causality: Understanding of problem and causal links between proposed inputs, outcomes and impacts</p>	<i>High</i>	<p>Complicated Rapid sequential learning with focus on how context influences problems and outcomes, for example through realist flavoured evaluation approaches</p>	<p>Simple Traditional approaches to performance monitoring using output indicators might suffice</p>
	<i>Low</i>	<p>Chaotic Multiple approaches to learning including real time monitoring and evaluation</p>	<p>Complex Parallel experiments & learning strategies with focus on identifying the most successful and identifying causal mechanisms.</p>
		<i>Low</i>	<i>High</i>
<p>Confidence in Context: Understanding of and ability to influence political context</p>			

Put simply, the diagram suggests that if those involved in programme design: (1) share an understanding of the problem; (2) have confidence that the solution proposed has a well-evidenced causal mechanism; (3) understand the context; and (4) believe it will work in the political and cultural context, then it will fit in the top right hand quadrant. A plan to inoculate child refugees accommodated in institutions in donor countries, for example, is likely to have predictable results on disease and could be evaluated by monitoring the number of children vaccinated. However, at the opposite end of the spectrum, there are interventions that belong in the more chaotic, left hand bottom quadrant. In these examples problems and contexts may be ill understood, or there may be disagreement about their characteristics. In addition the efficacy or causal mechanism of proposed solutions might also be unknown and thus a very different monitoring and evaluation approach would be necessary. Public health epidemics like Ebola and the Zika virus are obvious examples. They require a cautious trial and error approach in which real-time learning is not only directed at disease vectors and scientific medical interventions, but also at how political, cultural, religious and other

factors influence people's coping and mitigation strategies, as well as their engagement with any external intervention. Trying to apply blueprint standardised monitoring and evaluation approaches to such situations is unlikely to be possible or desirable. Those involved have to be innovative and flexible. Interventions located in the other two quadrants of the diagram are slightly less chaotic, but neither can rely on the RBM assumptions identified in the left hand side of Table 1. Hence methods for planning, monitoring and learning need careful consideration and might involve problem driven iterative adaptation that is outlined below.

Problem driven iterative adaptation – PDIA

Problem driven iterative adaptation – PDIA (Andrews *et al.*, 2012) is one of several management approaches promoted by the Doing Development Differently and Thinking and Working Politically communities that incorporates some of the complex adaptive thinking above.²⁸ Instead of assuming that actors have a mutual understanding of problems that can be tackled by proposing predefined technical solutions with predictable results, PDIA adapts ideas to contextually specific issues that are political in nature (Andrews *et al.*, 2012:6–7). Similar to process management used by DFID²⁹ and process consulting by SIDA³⁰ in the 1990s, PDIA was specifically designed to overcome weakness in the assumptions that underpin international development cooperation's externally designed, best practice solutions for institutional reform. Having established that 'best practice' solutions created incentives for senior state officials to appear to be implementing reforms, without influencing the kind of changes in political behaviours and norms necessary for civil servants to deliver services to poor people, Andrews *et al.* (2012) proposed an alternative approach.

PDIA is described as a locally led and politically smart process for identifying 'best fit' solutions. It involves four generic principles that

²⁸ Other approaches can be found in Hummelbrunner and Jones (2013).

²⁹ Mosse *et al.* (1998); Cracknell (1996).

³⁰ In the early 2000s Sida was using process consulting: "[W]hich means that the project starts from the definition of the problem but without a specific list of inputs and expected results. Specific goals, targets and inputs will be formulated as part of a process, the length of which is not normally determined at the outset." (Edgren, 2001: 60).

are applicable to a range of development interventions in national contexts:

- To define problems in particular ‘local’ contexts. The aim of the problem focus is to help actors deconstruct problems to get to root structural causes and for this to be the basis for coalition building and action plans. It is assumed this context analysis process requires a degree of political economy and power analysis.
- To attain approval from authorities to experiment with ‘small bets’ that are monitored through adherence to the third principle below.
- To create learning and feedback to identify challenges or positive deviants that can be further studied as local actors adapt.
- To engage the powerful and the powerless so that the opinions of the more marginal can be collected within processes that involve senior figures with power to influence decisions.

(Adapted from Hummulbrunner and Jones, 2013: 19)

4.2 Assumptions on knowledge, learning and evidence

Complexity and adaptive systems thinking has long informed the work of leading evaluation experts who have proposed evaluation designs that accommodate approaches like PDIA and avoid some of the weaknesses of evaluations underpinned by RBM. These include Sarah Earl (*et al.*, 2001) Michael Q. Patton (2011), Patricia Rogers (2008) and contributors to an edited volume on evaluating complex programmes and policies (Forss *et al.*, 2011). Some of these, for example Rogers (2008), incorporate theories of change. This is not only to enable project actors to move beyond linear assumptions of change in planning, but also to expose different stakeholder assumptions concerning the causes of problems and pathways to social change.

Patton (2011) and Reynolds (2015) use the concept of double loop learning made famous by Argyris and Schon (1974) to show that monitoring and evaluation plans driven by RBM logical frameworks are biased and seek to justify rather than inform policy decisions. They merely ask the question: ‘Are we doing things right to reach the target

or goal?³¹ Such a narrow line of questioning excludes possibilities for the kind of double loop learning that is necessary to answer the question: ‘Are we doing the right thing?’ Exploring this question requires a deeper examination of system dynamics to assess whether goals or strategies are the most appropriate courses of action. If evaluations (and the monitoring data that inform them) are to generate useful findings for policy, complex adaptive systems proponents argue they need to be driven by questions that set out to explore expected and unexpected results (Hummelbrunner, 2015:21; Bamberger *et al.*, 2016). Only then will it be possible to test assumptions about whether interventions were the most appropriate when compared with alternate goals and theories of change.

In addition to broadening the conceptual scope of evaluations, complexity informed approaches also indicate the need for a more contingent approach to overall monitoring, evaluation, learning and performance management approaches as suggested in Figure 1. If confidence in understanding and being able to manage risks in a context where an intervention that is known to ‘work’ is high, then standard RBM and performance management approaches could be appropriate and monitoring data may be sufficient to estimate results and outcomes. However, many development cooperation problems and solutions cannot be defined in advance. Therefore causal assumptions underpinning some proposed interventions need to be tested through an operational research, real-time or developmental evaluation that seeks to explore whether the right thing is being done or iterative adaptation is required (Ramalingham *et al.*, 2014).³² Information technology innovations such as crowd sourcing data using web-based platforms and mobile phones offer huge potential for real-time monitoring, adaptation and learning (UNDP, 2013).

According to participants at the recent USAID learning lab meeting where Figure 1 was developed, the precise approach to learning may involve parallel or consecutive experiments. Such approaches are equally applicable to global responses to challenges

³¹ I am grateful to Kim Forss for drawing my attention to the work of Bateson (1972) related to double loop learning that preceded Argyris and Schon’s (1974).

³² Developmental Evaluation is an approach that can assist social innovators develop social change initiatives in complex or uncertain environments. Originators liken their approach to the role of research & development in the private sector product development process because it facilitates real-time, or close to real-time, feedback to program staff thus facilitating a continuous development loop. Patton, M.Q. (2010). Developmental evaluation. Better Evaluation, [Online], Available: http://betterevaluation.org/plan/approach/developmental_evaluation [3 April 2016].

such as the refugee crisis, pandemics, climate change, terrorism and human trafficking as they are to more country specific problems. Patton is currently scoping the interest and possibilities of developing the capacity of Blue Marble Evaluators able to apply world systems analysis to global interventions.³³

It will come as no surprise to readers that critics of a short-term results focus argue for new approaches to monitoring and performance management that not only overcome assumptions that change is linear and predictable, but also that it is possible to generate generalisable knowledge about what will work in different contexts (Woolcock, 2013). According to Woolcock (2013), much 'learning' activity in international development cooperation is still driven by the assumption that it is possible to identify the best policy options through research or evaluation that identifies correlations between causes and average outcomes. But these methods, which include random control trials (RCTs) that are lauded for providing rigorous, randomised means to prove or disprove the links between intervention causes and effects, tend to be narrow in focus. Embedded in the same logic as RBM, they often fail to consider unexpected outcomes and the effects of the wider systems on results or the multiple perspectives of different actors experiencing them (Root *et al.*, 2015; Kleinfeld, 2015; Bamberger *et al.*, 2016). This reduces their predictive power in line with the assumptions related to the nature of evidence and knowledge outlined in the right hand column of Table 2.

Woolcock (2013) and Root *et al.* (2015) propose the use of case studies, grounded in theory-based realist thinking to overcome the risks of inappropriate evaluation designs outlined above. Realist approaches assume contexts – both geographical and personal – influence how and why innovations work or not (Westhorp, 2014). They seek to establish why certain groups of people engage with or respond to different initiatives in different ways. In other words, Woolcock's approach aims to assess the conditions under which diverse outcomes are achieved, paying particular attention to contextual and implementation peculiarities. Pritchett *et al.* (2012) also recommend structured approaches to the experiential evaluation of adaptive programmes by those implementing them. This is to enable real-time learning and adaptation. Such monitoring and

³³ Patton, M.Q. (2016). Blue Marble Evaluation, [Online], Available: www.utilization-focusedevaluation.org/blue-marble-evaluators/ [3 April 2016].

evaluation strategies provide possibilities to explore counterfactuals within programmes that undertake multiple simultaneous projects as part of their experimentation processes. It is assumed that comparing the relative effectiveness of different interventions aimed at the same problem in the same political, economic and social context will provide more relevant evidence for learning than evidence generated in completely different environments.

Proposals to integrate monitoring and evaluation processes to enable real-time learning are consistent with aspirations for RBM expressed by the evaluation community around the time of the Paris Declaration. However, Reynolds (2015) argues that the more recent focus on testing fundamental assumptions about *if* and how change happens in development interventions, which includes consideration of unexpected outcomes, is a departure from RBM's somewhat linear corrective logic.

Regardless of whether the MEL approaches proposed above enable single or double loop learning within a programme cycle, 'adaptation' that involves changing output and outcome level indicators during programme implementation poses challenges for results monitoring and performance management against organisational targets. To address this issue, those leading efforts to do development differently have come up with suggestions for adaptive indicators. These include performance metrics for monitoring the quality of learning and adaptation enabled by new MEL approaches,³⁴ and they are discussed in more detail in Sections 5 and 6 below.

Inclusive knowledge and learning

Participatory researchers and evaluators are not only bothered about which questions drive monitoring, evaluation and learning about results, they are also concerned by who asks the questions, what kinds of knowledge are considered valid and who learns, uses and controls the information. Chambers (2010, 2015), Guijt (2015), Wallace and Porter (2013) are likely to appreciate some of the realist approaches advocated by Woolcock (2013) and the intentions of Andrews *et al.* (2012) to include marginalised people in PDIA discussions. However, they may have suggestions about how such approaches could be made more participatory and useful for a range of local actors. Building on

³⁴ For example, see Mercy Corps (2015).

Holland's (2013) evidence of the potential of participatory statistics to define and assess results, Chambers (2015), for example, calls for a more inclusive and complexity informed definition of rigour that is able to address common methodological biases in policy research and evaluation. These include those that particularly disadvantage and exclude women from framing questions and telling their stories (Wallace and Porter, 2013).

Chambers' (2015) notion of rigour embraces methodological pluralism and acknowledges the need for researchers to improvise and innovate when dealing with complexity and uncertainty. While his suggestions are largely supportive and consistent with ideas found in PDIA, he also favours monitoring and evaluation approaches which address power relations that commonly influence research processes and findings. These involve including multiple perspectives in findings and celebrating inquisitiveness and ignorance on the part of assessors.

Many of the ideas above are synthesised in a useful paper entitled *Putting learning at the centre: Adaptive development programming in practice* (Valters *et al.*, 2016). The authors begin to explore how alternative approaches to development management can get better at engaging with the question of who, of the many stakeholders engaged in a development intervention need to learn what, when and where. The paper draws attention to the importance of focusing more on power, behaviours and relationships between actors involved in development cooperation, issues that have been somewhat absent in some conversations about complexity informed alternatives to RBM (Hummelbrunner and Jones, 2013; Mowles, 2014).

4.3 Assumptions about power, relationships, roles and behaviours

According to Eyben (2008, 2010), successful development cooperation has little to do with performance management tools used for project steering, it is all about the quality of relationships and the ability of individual staff to improvise and to take relational approaches to their work. Eyben (2005, 2006a, 2008, 2010) has consistently contested conceptualisations of development cooperation relationships in terms of RBM's principal agent theory. (Principal agent theory assumes that one party is made accountable to another through the use of contracts and performance management tools.) Instead she argues that the decisions of policymakers and practitioners

need to be informed by an appreciation of the pervasiveness of power in all relationships.

Results based management tools can have very real effects on practitioners and citizens in partner countries, and relational thinking provides a nuanced picture of how power operates through them. In the case of MEL staff mentioned earlier, who were empowered by the focus on results, effects can be positive (Whitty, 2015). Alternatively, reporting procedures can undermine trust and leave recipients feeling thoroughly disempowered (Causseman and Gohl, 2015; Wallace *et al.*, 2006). The kind of quantitative indicator data required in common results frameworks has proved detrimental for small women's organisations wanting to tell their stories (Wallace and Porter, 2013). But the effects of RBM tools go deeper than that. Abu Alghaib (2015:115) describes the excitement that a small, disabled people's organisation felt when a donor invited them to submit a proposal for funding. During the process of negotiations in which they were required to complete logical frameworks and adopt thinking from a donors' theory of change, however, they felt their "organisation became a stranger in its own project". Coming from a critical management position, Townsend *et al.* (2002) and Dar and Cooke (2008) argue the effects of RBM tools are more widespread and diffuse than Abu Alghaib's experience suggests. RBM tools also serve as an effective transmission belt for neoliberal ideology. They influence what is considered to be proper professional practice in countries that receive international development cooperation.

When it comes to overcoming power inequities in relationships between donors and partner NGOs, governments or supplier organisations caused by the RBM approaches mentioned above, ideas vary from the paradigmatic to the more practical. Chambers (2010) recommends shifts in rules and procedures as well as behaviours to transform performance monitoring and steering from policing and quality control activities to more empowering, mutual learning processes. His recommendations tend to focus more on the informal and personal attitudes and behaviours of individual staff than similar ideas put forward by Peace Direct and Mercy Corps, two of the NGOs advocating to influence donor management culture.³⁵ Developing adaptive approaches in partnership, increasing the accessibility of donor staff for informal conversations and developing means to value and assess more intangible outcomes that are poorly

³⁵ Mercy Corps. (2015).

described by standard quantitative results indicators are amongst the practical suggestions made by Mercy Corps (2015) and Peace Direct (Pinnington, 2014).

Encouraging practitioners to be more critically reflexive and to reflect on how power operates in aid relationships was at the heart of the Big Push Forward and of another similar initiative by gender activists (see Wallace and Porter, 2013). Both provided a space for practitioners to reflect on the politics of RBM and evidence paradigms. Participants were invited to share experiences of the visible and invisible power of results based discourse and tools in development cooperation relationships.³⁶ The BPF encouraged practitioners to reflect on assumptions about how RBM and the ideology that drives it shape knowledge, roles, behaviours and power relationships between those involved. Some in the development sector found this focus on politics too confrontational. Yet others have successfully adopted tactics identified by participants at the Politics of Evidence conference, facilitated by the Big Push Forward, such as taking collective action with donor staff to mitigate the most egregious effects of RBM. This is commented on in Section 5.

4.4 Summary

Critics of narrow interpretations of the results agenda agree on the need for planning, monitoring, evaluation and learning approaches to embrace new assumptions about the nature of problems that development cooperation agencies aim to address, the nature of change processes and the kinds of relationships required to achieve results. Some of them also draw attention to the benefits of development cooperation practitioners reflecting further on the nature of knowledge and evidence that enables learning and informs policymaking and practice. There is no shortage of conceptual ideas for tools informed by complexity science to enable a more politically smart, locally led and adaptive approach to planning, monitoring, evaluation and learning. Some of this thinking, particularly related to politics and power, has also informed recommendations for how to encourage more critically reflexive trust-based relationships. Selected evidence of the effectiveness of some of these approaches is outlined in the section below.

³⁶ See Eyben (2013) and the introduction of Wallace and Porter (2013) for details.

5. Emerging Examples of Alternative Approaches

Donors have found it difficult to move from thinking politically to working differently (Booth and Unsworth, 2014). But there are increasing examples that indicate they can do so and that an iterative approach to learning and problem solving that is unencumbered by results based management tools and other procedures enables success (*ibid*). In 2014, a review of seven ‘politically smart and locally led’ case studies, which had been independently evaluated or reviewed, found tangible evidence of short- or medium-term results that had benefited poor people (Annex 1). In each instance there was plausible evidence that the approach adopted was critical to such achievements. Reviewers found that the interventions were “demonstrably more effective than comparable efforts to address similar problems in similar circumstances; or ... pioneering in attempting to engage with the politics in a highly charged environment” (*ibid*: 7).

What follows is a critical discussion of recent examples of some of the alternative management approaches that seek to mitigate the effects of what the critics consider to be RBM’s most troublesome assumptions. They are mostly ordered as they relate to different stages in a typical programme cycle (bearing in mind, however, that some of these approaches defy a traditional programme cycle model). The examples come from a very small database of established cases and are not in any way presumed to be representative of the sector as a whole. Moreover, it is possible that unobserved some donor country offices are using alternative approaches while reporting to head office according to RBM requirements (Eyben 2010). The aim is to raise awareness of what is possible rather than reach any generalisable conclusions.

5.1 Political economy and power analysis

Advocates for Thinking and Working Politically argue that political economy analysis (PEA) plays a vital role in mitigating the negative effects of technical and reductive planning approaches encouraged by results based planning tools (O’Keefe *et al.*, 2014). They assume that more deliberate analysis of the distribution and contestation of power and resources within states, regimes and political parties will enable

learning with the potential to improve aid effectiveness (Yanguas and Hulme, 2015).

Political smartness was identified as a key factor responsible for the achievement of better than expected positive results in the small but growing number of case studies used by its advocates (Booth and Unsworth, 2014). However, despite individuals in agencies like DFID and Sida investing considerable energy in persuading colleagues of the potential benefits, both organisations have failed to achieve take-up at any scale (Fisher and Marquette, 2014; Yanguas and Hulme, 2015). PEA remains personality based and peripheral to core operational guidelines (*ibid*). Reasons given for this apparent failure include the macro-level and academic approach to PEA training (Fisher and Marquette, 2014; Hudson and Fisher, 2015), and organisational incentives that prioritise disbursement of funds and the achievement of quick results (Fisher and Marquette, 2014; Yanguas and Hulme, 2015). Consequently PEA is often a tick box exercise contracted out to external consultants (*ibid*), preventing any meaningful experiential learning by donor or partner staff.

In an article that exceptionally includes local ownership as an effectiveness criterion, Fisher and Marquette (2016) demonstrate how and why PEA risks being co-opted to serve donor domestic accountability agenda at the expense of encouraging trusting relationships and local learning. All too often PEA is undertaken by expatriate consultants and used for intelligence gathering and reducing potential exposure and risk. It neither involves nor is shared with partner governments or civil society.

Members of the Thinking and Working Politically community perceive an urgent need to overcome these issues of ownership and translate PEA into more practical knowledge for national actors. National stakeholders have a keen tacit understanding of local politics, though some argue they need inputs from external actors to contextualise and articulate it.³⁷ This review found several examples of efforts to advance such an agenda. Rather than relying on external consultants, the State Accountability and Voices Initiative in Nigeria, which brokers relationships between state, civil society and media, claims success in training and enabling local staff to make PEA part of their everyday practice. A recent paper by the Development Leadership Programme (DLP) provides a tool for those who have to

³⁷ This is a point that has been made in several discussions I have had with practitioners about local ownership.

make quick decisions in the wake of a political shock or change. It takes them through steps that encourage the consideration of interests that might be pushing agenda and reflection on the space and capacity they have to make change (Hudson *et al.*, 2016).

In order to be truly effective, political or power analysis has to include a gender lens, which has been somewhat lacking in the DDD and TWP discourse O'Neil (2016). This is either because those leading efforts do not see development as social transformation Green (2015), or because normative rights based approaches sit uncomfortably with locally led approaches embedded in gendered cultures (O'Neil, 2016). In a recent paper, O'Neil argues that the DDD and TWP communities have much to learn from the experience of feminist and women's movements' work on power and gender analysis. She also notes that those working on gender could gain from engaging with adaptive learning and management debates.

If political, power and gender analyses are to make any real difference they must be reflexive and include the influence of donors and other practitioners on interventions and outcomes (Eyben, 2003). Though rarely discussed in the literature, this is a point well made in the conclusions in a recent evaluation of the Ebola response. DuBois *et al.* (2015) argue that understanding the power relations and cultures of the development cooperation system and the communities it services is key to addressing the root causes of complex problems. Such thinking was central to the BPF's approach to encouraging reflection on the politics of evidence and results agenda.

5.2 Theories of social change

At the beginning of the decade some RBM critics were optimistic that theories of change would prove more effective than tools like the logical framework for encouraging practitioners and policymakers to debate their various assumptions about how change happens. It was hoped such discussions would encourage double loop learning (van Es and Guijt, 2015).

Although there is no single definition of a theory of change (TOC), it is generally understood as a means to communicate and explain implicit assumptions about how and why change happens in particular contexts (Stein and Valters, 2012). All policymakers and practitioners will have their own implicit assumptions about how social change happens that are influenced by disciplinary, cultural and

social backgrounds (Krznaric, 2007). Social activists saw an emphasis on TOCs providing an opportunity to encourage policymakers and practitioners to articulate and debate different understandings and assumptions of how history happens that would enable learning about different assumptions and the development of relationships. It was anticipated this would provide the possibility to move beyond the narrow range of theories of change from economic and political science that have tended to dominate international development cooperation's best practice development models (*ibid*). These include principal agent theory that has been popular because it is assumed to lead to quick results (Green, forthcoming; Krznaric, 2007).

Several international NGOs, for example Action Aid International and Oxfam International, have successfully used theories of change to get staff to explore their assumptions and come up with articulations like the example in Box 3.³⁸ Such broad articulations can then be made more specific in a particular context using locally led political economy and PDIA type power analysis to inform carefully monitored experiments.

As was the case with political economy analysis, there are few documented examples of this approach living up to expectations. Valters (2015) argues that TOCs have been co-opted to some extent through mandatory use in results based funding applications. According to some practitioners, demands for 'theories of change', which are essentially logic models used to describe how a particular project is assumed to work within a predefined donor theory of change, have resulted in them becoming 'logical frameworks on steroids' (Green, 2012). Yet despite such weaknesses there are anecdotal accounts of practitioners using theory of change assumptions as a basis for reflecting on monitoring data.

Box 3: Action Aid's Theory of Change

We believe that an end to poverty and injustice can be achieved through purposeful individual and collective action, led by the active agency of people living in poverty and supported by solidarity, credible rights-based alternatives and campaigns that address the structural causes and consequences of poverty.

People's Action to End Poverty

³⁸ ActionAid International. (2012). People's action to end poverty. http://www.actionaid.org/sites/files/actionaid/peoples_action_to_end_poverty_28pp.pdf [3 April 2016].

SAVI and the Asia Foundation have both found theories of change a useful basis for learning and adaptation decisions during programme implementation (Ladner, 2015; SAVI, 2014). Indeed, there are a few examples of programmes dropping unsuccessful components as a consequence of such activities. However, while recent reviews suggest theories of change are proving useful tools for single loop learning, there appear to be fewer examples of them being used to ask the kind of radical and searching double loop learning questions that actors advocating for an alternative management paradigm might expect (Valters, 2015).

Literature and blog debates on theories of change suggest they have bothered, bewitched and bewildered many practitioners and this appears partly due to their quite distinct origins and uses (Green, forthcoming). Social activists who believe change is complex and unpredictable view theories of change as useful tools for exploring assumptions about how history happens in a particular context. Evaluators, on the other hand, are more interested in testing whether such assumptions can be causally verified in programmes that are trying to accelerate such change (*ibid*). Related, practitioners are often confused about how to link different kinds evidence with a theory of change approach (Stein and Valters, 2012; Valters, 2015). In instances where there is no established evidence base for a particular intervention causing a result, they question whether programme managers should base their theory of change on evidence or social theory from other contexts as opposed to their more experiential knowledge of how change tends to happen in their contexts (*ibid*).

Embedding theories of change in RBM procedures limits their potential for double loop learning, but there are other obstacles to engaging in deep discussions about how change happens. Within the context of ‘messy partnerships’ (Guijt, 2008), comprised of people with different understandings and opinions, including about the quality of data used, such questioning can be emotionally as well as intellectually challenging (Eyben *et al.*, 2008). A review of a multi-country governance and rights programme suggested that staff’s simplistic assumptions about governments responding to citizens’ rights claims appeared reasonable in some countries. However, in others like Angola, deep questioning led some staff to question whether the pursuit of a western democratic governance model was

the most appropriate development cooperation goal.³⁹ Exploring the possible implications of such reflections was difficult and troubling.

Despite it proving difficult to challenge theories of change and deeply held beliefs in formal programme settings, such questioning can and does happen. Both the DDD and TWP communities of practice are founded on the ‘lesson’ that principal agent assumptions underpinning institutional reform programmes supported by development cooperation rarely work (Booth, 2012a, 2013). Although state institutions appear to implement reforms, these have had little effect on frontline performance (Andrews *et al.*, 2012). Programmes like the State Accountability and Voice Initiative in Nigeria, SAVI, are therefore experimenting with the assumption that collective action or changes in informal institutions, such as social norms, offer a more effective alternative for achieving results than other theories of change.

5.3 Monitoring, evaluation and learning

Current debates about adaptive programmes are leading to innovations that appear to have enabled politically smart, locally led flexible programmes to learn and adapt while also meeting performance monitoring data requirements. However, it is not always evident if performance-monitoring data is supporting that learning. One case, an evolutionary experiment using positive deviance⁴⁰ by Oxfam in Tanzania (Green, forthcoming), uses approaches that resemble some of those promoted by Woolcock (2013), mentioned earlier. It demonstrates the principle that ‘small bet’ experiments and participatory monitoring by community actors can help to identify positive deviation by groups that find better solutions to problems than their peers. Oxfam supported several different groups to increase citizen participation on various issues such as school management. Then, in line with real-time learning and adaptation, communities, partners and Oxfam chose a date when staff came together to monitor progress and identify the most successful variants. In this case a group of farmer activators had enjoyed the best results and as a consequence they were supported to refine and expand their work. Green

³⁹ Personal experience during a Mid-term Review in 2013

⁴⁰ More details on positive deviance can be found here <http://www.positivedeviance.org> [3 April 2016].

(forthcoming) claims initial results were impressive. However, he does not elaborate on whether the data used for learning and adaptation was generated through processes that ran parallel to other procedures for measuring and managing performance using standard quantitative results data. The risks of parallel processes need to be considered as is evident in the report of an ‘outcome harvesting’ pilot project by the World Bank.

Like many other agencies, the World Bank has been piloting outcome harvesting as a means to learn from and adapt its knowledge support programmes in complex contexts (Gold *et al.*, 2014). Outcome harvesting ‘harvests’ evidence on what has changed – outcomes – and then works backwards to determine whether and how an intervention has contributed to these changes.⁴¹ A review of cases in ten countries enabled learning about expected and unexpected outcomes, particularly changes in the attitudes and behaviours of actors, and how these were achieved. But this ‘results learning’ generated data that complemented rather than substituted data from random control trials or other monitoring and evaluation approaches. The results data appeared to be used in the outcome harvesting process, but the extent of its utility was not entirely clear in the report that says little about whether the programme was constrained or helped by RBM and performance management approaches during early implementation. The review authors conclude that there is a need to build capacity to capture outcome data more regularly and integrate and analyse it with other standard data and include analysis in regular reports. It is hard to assess what impact such recommendations might have and whether they risk outcome harvesting, like PEA, becoming another bureaucratic requirement with all the standard problems associated with RBM tools.

Other programmes seem to have fared better in breaking free from the perceived constraints of traditional RBM strictures, such as quantitative indicators used for performance monitoring, that provide limited information on what a programme is doing or achieving. In a recent paper by the Asia Foundation, Ladner (2015) concludes that staff engaged in a programme funded by Australia’s Department of Foreign Affairs and Trade (DFAT) have found regular structured reflections using a monitoring process called ‘strategy testing’ useful for adaptation. Described as an alternative to traditional methods that

⁴¹ More details on outcome harvesting available Better Evaluation: http://betterevaluation.org/plan/approach/outcome_harvesting. [4 April 2016].

track progress against pre-determined outcomes, there are no baselines at the outset of the programmes. Outcomes are defined and refined once outputs start emerging and strategies take shape during the process of project implementation. It has undergone four rounds of testing in nine countries and included a range of the Foundation's programmes, from improving urban services in Mongolia to policy reform in Bangladesh. Similar monitoring and learning approaches were used by three of the seven programmes reviewed by Booth and Unsworth (2014).

In another example Van Ongevalle *et al.* (2014) document a complexity informed action research project that explored different actor-focused planning, monitoring and evaluation approaches. Like outcome harvesting and outcome mapping, which has been used by Sida, actor-focused approaches emphasise monitoring and learning from changes in attitudes and behaviours. The action research project enabled the ten participating NGOs and their southern partners to develop an analytical framework to help organisations explore processes of social change with unanticipated results that are difficult to measure or to quantify. This case, similar to others cited above, broadly reflects the experiences of monitoring and evaluation expert members of the DDD manifesto community who have successfully piloted flexible approaches to monitoring and evaluating results. Lessons resulting from their experiences that were shared during a 2014 workshop can be found in Box 4 below.⁴²

⁴² IDL Group GRM (2014). Monitoring and learning in politically smart and adaptive programmes. Workshop Note. 11 December 2014; Additional comments can be found at Barr, J. (2015). Monitoring and evaluating flexible and adaptive programming. Itad, [Online], Available: <http://www.itad.com/monitoring-and-evaluating-flexible-and-adaptive-programming/> [3 April 2016].

Box 4: Adaptive Management: Attributes of Successful MEL Systems

- Logical frameworks that are either loose enough to allow adequate room for manoeuvre, or are regularly updated, especially at lower (output and below) levels
- Getting staff involved in setting and adapting indicators so they extend the adaptive management principle to MEL systems, whilst also building local ownership
- Baskets of indicators that include some bedrocks that do not change for donor reporting; flexible output indicators that can be dropped as the programme evolves; and concrete change indicators that are unidentified at the beginning of the programme and then stipulated at a later date
- Theories of Change that provide evaluators, managers, and frontline implementers, with a regular means to review and revise their approaches. Good TOCs also provide a shared language and framework that enables joint learning and strategising with partners
- Documenting contributions to policy during programme implementation to enhance evaluability
- MEL being an on-going and in-house process (with external support if necessary) to enable tighter feedback loops between monitoring and changing strategic and operational tactics
- Political savvy, with senior managers buffering programme staff from donor demands and mechanistic reporting requirements by finding creative ways to deliver evidence and results to their funding bodies
- Using tools and approaches, such as stories of change, outcome mapping, reporting failure, politically/power informed context analysis

(Based on analysis of an IDL GRM December 11 2014 workshop note and SAVI's M&E Approach Document)

Even if some programmes mentioned above use log frames that include quantitative data of limited utility, donor staff and programme

managers have successfully innovated and found ways to mitigate some of the log frame's most deleterious effects. SAVI, the State Accountability and Voice Initiative in Nigeria, also recognised as a 'politically smart, locally-led' success story (Booth and Chambers, 2014),⁴³ deserves special mention.

SAVI is an 'arms length' programme that claims an identity as a broker of relationships between actors from different sectors, rather than being a funder or source of money. What is particularly remarkable is its approach to monitoring, evaluation and learning. Local staff have been involved in the development of what appears superficially to be quite a standard RBM performance management framework; however, it includes innovative aspects that have enabled flexibility and learning. SAVI successfully used a basket of indicators including a few that were clearly defined at the start of the programme and others that were 'filled in' as the programme learned more about the context. Retrospective outcome harvesting and stakeholder perception scores of changes in state level outcomes were two methods used to measure results. SAVI claims that this approach enabled the programme to provide DFID with the results data it needed without frustrating learning. Adaptations made as a result of learning and failure have been documented; the examples provided are convincing in terms of demonstrating that monitoring and learning from results systems can work to support adaptive development management.⁴⁴ Nevertheless, the investment in relationships needed to do this, which some donors label as 'transaction costs', were high. It is well documented that individual DFID staff played critical roles in supporting an adaptive approach that enabled the programme to meet the performance monitoring reporting requirements of DFID Headquarters (Booth and Chambers, 2014).⁴⁵ By August 2015, the SAVI results framework was in its fourteenth official iteration since the programme started in 2008.

⁴³ The description of the framework comes from a document on SAVI's M&E approach. SAVI. (2015). Defining and measuring results: monitoring incremental change over time. Approach Paper. SAVI, [Online], Available: http://savi-nigeria.org/wp-content/uploads/2015/12/SAVI_ApproachPaper5_2015_FINAL.pdf [3 April 2016].

⁴⁴ SAVI. (2014). Rising to the challenge: supporting 'problem driven iterative adaptation' and 'politically smart, locally led' approaches through a donor-funded programme The experience of the State Accountability and Voice Initiative in Nigeria http://savi-nigeria.org/wp-content/uploads/2014/10/DFID_SAVI_brief_Challenge_Online.pdf [3 April 2016].

⁴⁵ Personal communication from SAVI MELconsultants.

Efforts to use experience from the examples above to shift norms and mainstream learning driven approaches to M&E have gathered momentum in recent months. Staff from the Overseas Development Institute have been working with DFID country programmes to put ideas from Valters *et al.* (2016) into practice. Another example is the focus on learning and adaptation in a recent meeting of practitioners belonging to the World Bank's Global Partnership for Social Accountability.⁴⁶

Despite the progress noted in this section, old debates regarding relationships between monitoring and evaluation and what kinds of learning and evidence are good enough for decision making mean challenges still lie ahead. In a recent World Bank paper Gonsalvez De Asis and Woolcock (2015) argue that the Bank needs to abandon technical best practice models. They recommend the development of integrated MEL strategies that are based on Doing Development Differently and Thinking and Working Politically assumptions about the political nature of uncertain problems as well as pathways to change instead. However, marketing such ideas in multilateral institutions is challenging. Additionally, deciding exactly what kind of evidence and learning opportunities different actors need at different times is a formidable task. Woolcock (2013) and Gonsalvez De Asis and Woolcock (2015) show that generating an evidence base for adaptive programmes is likely to involve methodological and political arguments concerning whether donors can or should seek meta-theories for the replication of complex programmes in different contexts.

Achieving consensus in large institutions like the World Bank about what evidence is good enough to support adaptation will require heated debate between people with strong beliefs and values about the nature of evidence.⁴⁷ As of now the debate within big institutions does not seem to have included much consideration of whether and how to increase local participation in MEL processes. Obtaining feedback from citizens is deemed important by those advocating for change at the World Bank (Gonsalvez De Asis and Woolcock, 2015). But, if

⁴⁶ Moses, M. (2016). Trying, Learning, Adapting at the GPSA Global Partners Forum. Global Integrity: Data, learning and action in open governance, [Online], Available: <http://www.globalintegrity.org/2016/05/trying-learning-adapting-gpsa-global-partners-forum/> [20 May 2016].

⁴⁷ Woolcock's paper and other encounters with World Bank Evaluation staff reveal tensions between those who view random control trials and cost efficiency analysis as 'gold standards' and are not keen to entertain alternative methods.

methods used do not involve citizens in framing questions or data analysis, they can prove extractive. On the whole, meaningful discussions of the potential benefits of more participatory approaches within the DDD and TWP communities appear in a relatively nascent stage. This may change as engagement with those working on adaptive approaches in the NGO sector deepens.

Admittedly, broad participation of citizens in MEL processes may not always be appropriate or possible, but locally led evaluation is fitting with the overall principles of Doing Development Differently and Thinking and Working Politically. Evidence from *Local First in Practice: Unlocking the power to get things done*, a study funded by Sida (Pinnington, 2014), emphasises its benefits and the possibilities provided by the spread of information communication technology mentioned earlier. In an unrelated example Shah (2014) presents two evaluation case studies that demonstrate the potential of the most significant change (MSC) narrative technique. MSC enabled an evaluation that created space for a more participatory account of a programme's function. It succeeded in capturing unexpected outcomes, while also acting as a tool for real-time formative learning. What is particularly interesting is that the cases expose the competing theories, logics and values behind programmes while also legitimising the use of local programme knowledge for assessing the value and worth of the programmes.

Furthermore, a considerable evidence base demonstrates the potential of participatory approaches to generate robust statistics for research, monitoring and evaluation (Holland, 2013). One frequently cited example is the *Measuring Empowerment? Ask them* study, also supported by Sida. It enabled thousands of poor people in Bangladesh to measure progress against their own indicators in ways that were judged to be meaningful and fair.⁴⁸

Participatory approaches often raise questions related to bias, feasibility and rigour. These are teased out in methodological reflections on a large Participatory Impact Assessment and Learning Approach (PIALA) to assess the impacts on rural poverty of two government programmes in Vietnam and Ghana funded by the International Fund for Agricultural Development (IFAD) (Van Hemelrijck and Guijt, 2016). The authors conclude that there are trade-offs between bias, feasibility and rigour in all evaluations that can be reduced by out-of-the box thinking and the building of

⁴⁸ Jupp et al. (2010).

sufficient research and learning capacity. More generally though, methodological issues related to biases induced by incentives in the international development cooperation system have not been systematically addressed in debates about evaluation amongst Thinking and Working Politically and Doing Development Differently communities (Camfield *et al.*, 2014; Picciotto, 2016).

5.4 Relationships, roles and behaviours

Assessing the extent to which specific recommendations about enhancements to results based tools and associated behaviour change has impacted on the quality of development cooperation relationships is not easy in a desk-based review. Unsurprisingly, personal relationships have been somewhat ignored until recently in debates about results based management and political and adaptive programming (Fisher and Marquette, 2016). This seems to be changing.

Booth and Chambers (2014) talk specifically about individual DFID staff who made a difference when it came to managing relationships with SAVI managers. Informal interviews with practitioners working on adaptive programmes endorse the view that, confident and politically savvy donor staff who are able to navigate the results based frameworks and contracting procedures to enable flexibility make all the difference.⁴⁹ Honig (2014) provides evidence of the correlation between the discretion staff have in complex contexts and development results. This formal acknowledgment of the important roles individual staff and relational skills play in enabling politically smart programmes to achieve results is promising. However, the later discussion on human resource policies in Section 7 highlight the multi-causal factors that influence the practice of negotiating performance targets and managing relationships. They suggest it is likely to be difficult to change power relations in development cooperation without a more significant shift in the management paradigm. Reworking a results framework fourteen times requires significant staff effort, which is likely to be challenging in donor agencies, like DFID, that are working within a public sector model that values extreme efficiency. A recent publication engaging with those advocating complexity informed approaches to adaptive

⁴⁹ Interviews with Pete Vowles and another Key Informant.

management calls for realism. The author notes that some of the assumptions relating to the quality of relationships are overly optimistic and unlikely to hold true in programmes still operating in risk-averse, results driven environments (Buffardi, 2016).

5.5 Payment by results

Payment by results (PBR) or results based financing is a relatively new aid modality. It divides opinion on whether it encapsulates all of the problems of a domestic facing accountability agenda or whether it provides an important new tool that can support locally owned development. Sida seems optimistic about its potential (Sida, 2015), as were several influential advocates for DDD and a more politically informed approach to development (e.g. Andrews *et al.*, 2012; Barder, 2012b; Glennie *et al.*, 2013; Ramalingham, 2013). Their initial support was based on the normative premise that PBR would help partners and suppliers agree payment milestones in terms of outcomes, without donors advising on approaches or solutions to be used.

Exchanges between UK aid policymakers and commentators at the Center for Global Development suggest that influential actors have different theories of the potential benefits of payment by results type models.⁵⁰ Barder uses a complexity lens to argue that the main advantage of payment by results should be flexibility and local ownership for partners.⁵¹ However, Stefan Dercon at DFID uses principal agent theory to emphasise the benefits of PBR as a tool to incentivise behaviour change in service providers. This suggests quite a different kind of relationship and one that Eyben and Guijt (2015) argue risks instrumentalising rights-based approaches. For example, education interventions start to focus on the ‘what’ of PBR metrics that can be measured like enrolment rates instead of the ‘how’ of education being a right to engage in a process of empowerment.

In practice, PBR is not a one size fits all model. It requires considerable fine-tuning to accommodate the idiosyncrasies of each

⁵⁰ A theory based review of an NGO multi-country democracy and human rights programme provided a useful opportunity to identify missing assumptions about how a programme was going to contribute to gender equality.

⁵¹ Barder, O., Perakis, R., Savedoff, W. and Talbot, T. (2014). 12 principles for payment by results (PbR) in the real world. Center for Global Development, Views from the Center, [Online], Available: <http://www.cgdev.org/blog/12-principles-payment-results-pbr-real-world-0> [4 April 2016].

programme to get it right.⁵² A recent discussion amongst practitioners who had conducted literature reviews of payment for results type programmes or had direct experience of payment for results based mechanisms confirmed that PBR modalities can work.⁵³ But the conditions under which they work are highly contingent on the context, theory of change and level of indicators set as payment milestones. There are examples of badly designed programmes proving expensive for NGOs, but also examples of well-designed programmes having positive results. For example, an international programme working on reducing the use of carbon energy supported by DFID was reported to be having positive effects on the kerosene market in Tanzania.⁵⁴ This appeared due to the programme's PBR design providing incentives to firms that held the government to account. However, consistent with findings by Vahämäki, *et al.* (2011), most commentators challenged the assumption that finance was the lever to change behaviour. They gave examples of the publication of performance data related to payment milestones having more impact on service delivery through peer pressure mechanisms than the payments themselves.

Given confusion about the theory of change underpinning PBR models, as well as uncertainty about who is going to finance the costs of learning from failure,⁵⁵ the chances of the model providing an opportunity to support long-term adaptive programming informed by complexity science assumptions appear remote. Though its application to politically savvy designs and nuanced theories of change seems to produce promising effects, these may be rare. Other examples suggest that it is often delivered in line with principal agent theory and best practice models designed to achieve quick, low-level results (Vahämäki *et al.*, 2011). No one involved in the discussion about PBR mentioned

⁵² Barder, O., Perakis, R., Savedoff, W. and Talbot, T. (2014).

⁵³ Comments from participants in discussion on Green, D. (2016). Payment by results hasn't produced much in the way of results but aid donors are doing it anyway – why? From poverty to power, [Online], Available: <http://oxfamblogs.org/fp2p/payment-by-results-hasnt-produced-much-in-the-way-of-results-but-aid-donors-are-doing-it-anyway-why/> [4 April 2016]. They included Jake Allen of Itad who had just completed a review for the OECD, Donald Menzies a DFID PBR Policy Staff, Michael O'Donnel Head of Learning and Effectiveness at BOND the NGO Network and Stuart Worsley the co author of a recent book on Complexity <http://oxfamblogs.org/fp2p/payment-by-results-hasnt-produced-much-in-the-way-of-results-but-aid-donors-are-doing-it-anyway-why/> [3 April 2016].

⁵⁴ DFID. (2015). Annual Review Report Sheet for Results-Based Financing for Low Carbon Energy. DFID, [Online], available: bit.ly/1T6nRuG [3 April 2016].

⁵⁵ Green, D. (2016).

above described the kind of hands-off model that enables learning and adaptation outlined earlier. In fact, several examples suggested quite different donor behaviour. A few commentators remarked that DFID still insisted on the micromanagement of portions of payment by results project budgets.

Positive effects reported included PBR creating a space to discuss performance and also leading to improvements in M&E capacity for quite low-level performance indicators, but these did not necessarily enhance local level learning. Moreover, there are fears that instead of supporting the development of country level M&E systems, PBR may lead to parallel systems controlled by donors. These would increase upward rather than downward accountability of partner governments to their citizens. An interesting example cited that challenged this idea came from an NGO PBR validation exercise in which communities reported that the project involving a PBR mechanism had increased the NGO's accountability to them. Other NGOs, however, noted that if payment by results become the norm it could fundamentally change their relationships with communities. Instead of being viewed as solidarity partners there was a risk they would be seen as nothing more than service providers of transactional development.

5.6 Summary

There are increasing examples of politically smart planning, monitoring, evaluation and learning approaches promoted by different groups challenging the negative effects of a short-term results focus proving effective. Yet the cases reported are still few and far between. They almost certainly under represent the efforts of a large number of practitioners who have always worked in these ways, but perhaps are operating under the radar.⁵⁶ One important contribution the DDD and TWP communities have made is highlighting the importance of these actors and their skills in managing messy power relationships.

Currently it appears that demonstrating the efficacy of alternative management approaches to donors is a priority, and this poses risks. Notwithstanding the progress made, efforts to enhance results through political economy analysis and theories of change have not

⁵⁶ Several people have remarked that the approaches that are now getting attention because of the push back against DFID's rigid approach to results planning implemented in 2010 describe the standard practices of seasoned development practitioners.

lived up to expectations and efforts to incorporate gender issues and critical reflexivity on the role of development agencies are disappointing. Moreover, efforts to advocate pluralist evaluation approaches are important and reflect awareness of different assumptions of the nature of knowledge, but they probably need to go slightly further to satisfy participatory practitioners like Robert Chambers and Tina Wallace who have been critical of RBM's effects on the just and fair use of knowledge for years.

Furthermore, the few examples of success cited here raise age old questions about whether performance monitoring systems are going to run in parallel with learning systems, generating data that is only useful for political accountability and for managers or researchers interested in performance management. Early findings on payment by results programmes suggest this might be the case; the likelihood of the model providing an easy means to increase local ownership and build the capacity of partner country M&E systems seems unlikely. Despite development cooperation appearing to abandon principal agent theory in complex contexts, it still provides a default setting for aid relationships, possibly limiting what individual staff pushing for change can be expected to achieve.

6. Implications for Organisational and Institutional Change⁵⁷

Those advocating for more flexible approaches to development management are mindful that mainstreaming efforts will require non-trivial organisational change. This section considers their suggestions that donors will need to reconsider: policymaking processes; the design of results based management and MEL systems; finance, procurement and contracting policies; human resource policies and procedures; and public communication messaging. It draws attention to the relevance of earlier discussions about the problematic assumptions underpinning new public management inspired RBM tools for organisational change.

Many organisational performance management systems are premised on assumptions that donor agencies operate as bounded bureaucratic machines and that staff practice is determined by rules and financial incentives. Yet examples of individuals from donor and partner agencies innovating and influencing donor performance management, including through the DDD and TWP communities, indicate principal agent theory does not provide a full picture. In practice donor agencies are porous and the behaviour of individual staff members is contingent on their informal relationships and the norms of their wider networks. Thus many of the changes discussed in this section are better informed by views of organisations as contingent and unpredictable entities, rather than rigid bureaucracies (Gulrajani and Honig, 2016).

6.1 Policymaking

None of the RBM critiques takes issue with the notion that programmes should be informed by reasonable evidence, however, Woolcock (2013), and Root *et al.* (2015) cited earlier highlight the challenges that adaptive programmes could present to ‘best practice’ policymaking traditions. There is considerable debate about the

⁵⁷ This section draws on literature review findings and my personal knowledge and experience. Although some points are linked to issues discussed in other sections, they are not derived from previous sections as would be the case in a more empirically grounded study.

relevance of different kinds of evidence for policy and programming in different contexts and at different points in programme cycles (Gonsalvez De Asis and Woolcock, 2015). Yet a recent report on the failings of the Ebola response that prioritised technical and scientific evidence over understanding of the local political context (Dubois *et al.*, 2015) illustrates how costly mistakes can be. It cautions policymakers to resist any call for universally relevant evidence of ‘what works’.

Uncertainty about what kinds of evidence are most appropriate to use at different times, together with the inherent methodological biases associated with most normative social science research methods (Morton, 2009; Camfield *et al.*, 2014), make good policymaking extremely difficult. These factors indicate that both policymakers *and* those trying to influence policy inevitably have to make do with less than perfect evidence and ‘muddle through’. Thus they will need to be reflexively conscious of the fragility of evidence and the threats of bias that are exaggerated by the political economy of development cooperation (Morton, 2009; Camfield *et al.*, 2014). The risks of harm resulting from well-intentioned development entrepreneurs overselling ideas are very real. Questions are currently being raised about whether payment by results financing will end up being one such example for the reasons provided above.⁵⁸

6.2 Monitoring, evaluation, learning and performance management systems

Although there is variation between donors, to date their results based management systems have not been wholly effective in supporting different results agenda (Gonsalvez De Asis and Woolcock, 2015; Lloyd *et al.*, 2014; Manning and White, 2014; OECD, 2013, 2014, 2015). While enabling some donors to appear effective and functionally accountable, findings presented earlier raise questions about: (1) the reliability of results monitoring data for assessing effectiveness without the integration of randomised impact evaluation data (Manning and White, 2014); and (2) whether decontextualised donor centric output-level results data could be useful for learning at an aggregate global level.

⁵⁸ Green, D. (2016).

Wild *et al.* (2015) argue it is time for a new approach to measuring performance and results that can enable real-time learning, for example through selecting performance indicators to assess the quality of learning and adaptation to different contexts, and measurements of innovation. Their argument is that this would help to generate more useful evidence for learning than current output-level data and therefore help achieve better results overall.

Suggestions for basing performance indicators on learning are logical and useful in as much as they try to increase strategic accountability. In addition if implemented they could help to integrate results data with learning that is currently maintained in other systems in more systematic ways. However, experience raises questions about if and how such approaches could be effective without greater investment in the development of trusting relationships which would require more staff than most donor agencies have today. Additionally, early evidence emerging from PBR contracts suggests that the donor community is going to struggle to move away from an emphasis on quantitative output indicators. Hence there is a real chance that any new 'learning metrics' will end up being monitored in parallel with donor systems for tracking output data. Furthermore, if, as the Big Push Forward argued, the problem is the RBM logic and not the indicators themselves, then it is also important to reflect on whether changing the goal posts will influence the rules of the game.

Whichever way one looks at it, developing single management information systems that are able, on the one hand, to ensure adequate quality programmes are in the pipeline to be confident of achieving results targets, while on the other hand, enabling adaptive programmes to accommodate some space for learning from failure at outcome level is not straightforward. And, even if definitions of local ownership are shifting in the new politically smart, locally led, complexity aware paradigm, questions still have to be asked about how donor actions will enable the development of partner government systems. Partner governments and other local actors will need support if they are going to be able to track progress against their development targets and the new Sustainable Development Goals (SDGs). According to some like Manning and White (2014), donor contributions to the SDGs should be a primary measure of donor performance.

Clarifying assumptions about who needs to learn and for what is obviously a central question for those designing new management systems. Donors are always expected to learn yet recent reports by OECD (2014: 20) and ICAI (2014) drawing on experience from

DFID suggest learning expectations need to be realistic. Staff in DFID were reported to be overwhelmed by the sheer volume of intellectual products they were expected to digest. If (1) the external validity and thus utility of de-contextualised impact evaluation is in question (Woolcock, 2013); (2) donor staff numbers are going to remain static or decrease; and (3) complexity and politically informed programming requires more investment in local real-time learning, then perhaps it is time to revisit assumptions about what learning should take place in donor headquarters. Blue Marble Evaluation arguments that highlight the need for global learning agenda that relate to global challenges like climate change provide the opportunity to re-imagine the kinds of systems donors need for performance management.⁵⁹ They also provide space to reconsider the relationships between monitoring, evaluation and learning and the capacities likely to be needed to perform these functions in different geographical locations.

Politically smart complex programmes, including those responding to global challenges, require different kinds of evaluation designs. This has implications for commissioning and delivery capacity in donor and partner countries. Donor agencies and partners need to be able to commission evaluations using mixed designs, for example including realist and case study approaches, ethnographic and participatory methods (Stern *et al.*, 2012). This will not be easy given the challenges these methodologies present.

6.3 Financing and procurement

Conversations with practitioners involved in adaptive programmes suggest the need for both minor and more radical changes to financial planning and value for money considerations. Hummelbrunner and Jones (2013) argue that an experimental approach to programming requires less money be spent on initial detailed planning of interventions, though it could be argued that there is a need for a greater investment in developing an overall monitoring and evaluation plan early on in programmes (Shutt and McGee, 2013).

Other cost relationship changes implied by arguments made by the Doing Development Differently and Thinking and Working Politically communities are likely to have more radical implications for

⁵⁹ Patton, M.Q. (2016). Blue Marble Evaluation, [Online], Available: www.utilization-focusedevaluation.org/blue-marble-evaluators/ [3 April 2016].

how development cooperation agencies think about efficiency and value for money. Reducing the amounts of money offered to partners while also demanding significant donor and supplier investment in relationships and the constant renegotiating of performance measures promises to increase what have traditionally been considered ‘transaction costs’. One key informant involved in such a programme raised questions about whether it would be possible to scale up such an approach in the current economic paradigm that often seems to value efficiency (low administration or programme support costs) over effectiveness (the costs of relationships). Moreover, finding ways to cost in failure that will reduce benefit cost ratios poses challenges to traditional cost–benefit analysis approaches to making value for money judgements. These kinds of cost relationships would, for example, call into question some of the thinking underpinning DFID’s value for money strategy that mainly focused on efficiency and economy metrics rather than sustainability and impact (ICAI, 2015). Impending discussions about how to design and implement effective PBR models provide important opportunities for conversations about financing approaches that should also include discussions of contracting models. Ideally these would be informed by a view of donor agencies being contingent rather than bureaucratic organisations.

Policymakers and practitioners advocating for more politically, relationally smart adaptive programming are acutely aware that performance measurement systems that place huge emphasis on correct financial forecasting will make much of what is proposed hard. As these systems are driven by government cash management requirements they tend to be inflexible. Hence, such difficulties are not easily overcome.

6.4 Decentralisation and human resource policies

Giving staff more discretion in the use of rules and procedures for performance management was a key motivation behind the development of what are referred to as DFID’s new ‘SMART’ rules.⁶⁰ Following some of the critical reports about the effects of DFID’s

⁶⁰ DFID. (2016). SMART rules: better programme delivery. UK Government, [Online], Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/472771/Smart_Rules_1November_to_31_March_2016.pdf [3 April 2016].

approach to RBM discussed in Section 3, DFID staff members pushed for a change in rules and procedures that would enable more flexible programming and performance management approaches. The rationale for these SMART rules, which has been promoted by experts like Barder (2014), is supported both by anecdotal and, seemingly, robust evidence. This evidence suggests there is a relationship between individual managers' discretion and poverty reduction results, particularly in complex contexts (de Renzio *et al.*, 2005; Honig, 2014).

These changes notwithstanding, it is likely to take more than a tweak in performance management rules to alter behaviour and to empower donor staff to use more discretion in relationships with partners. Although the move to DFID's new rules was well received (ICAI, 2014), it did not have immediate effects. Policy staff and corporate managers based in London found it difficult to let go of trying to control front-line operations. This is despite organisational knowledge that central decisions have limited effects on front-line operations.⁶¹

Furthermore, despite a significant reduction in mandatory procedures, such as the use of the logical framework, few DFID staff or partners have yet come up with alternatives.⁶² One possible explanation is that the change was ambiguously articulated in the new 'SMART' rules; another is Lipsky's (1980) argument that official rules have limited impact on the behaviour of front-line civil servants.⁶³ What seems more likely, however, is that one technology of control was replaced by another, which is a typical feature of NPM. Though the new 'SMART' rules were meant to decentralise power to country office staff and provide them with more freedom in the application of rules (DFID, 2016), this has been accompanied by another change that has increased personal responsibility for programmes. The managerial language in job descriptions for Senior Responsible Owners that makes programme managers 'owners', who are responsible for "delivering objectives" and monitoring to "take control of progress"⁶⁴ may override the flexibility intended by the

⁶¹ Pete Vowles' comment on Barder, O. (2014).

⁶² This was corroborated in an interview with Pete Vowles, DFID's Deputy Head of Programme Delivery, 4 December 2015.

⁶³ This was corroborated in an interview with Pete Vowles, DFID's Deputy Head of Programme Delivery, 4 December 2015.

⁶⁴ Taken from sample SRO letter of appointment Lowcock, M. and Blackall, D. (2015) Letter of Appointment, UK Government, [Online], Available:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/415697/Letter-of-appointment-SRO-StHelena-airport.pdf [3 April 2016].

SMART rule change.⁶⁵ Given the tone of the language staff located at country level may become even more anxious and controlling. While there is no specific evidence to prove this particular causal link, the broader discussion supports the more important point – that changes in advice about the use of performance monitoring and management tools may have limited effects on trust, relationships and practices consistent with an alternative development management paradigm. Hence, contingency theories about the relationships between leadership, human resource policies, performance management approaches, trust and organisational environments deserve more policy attention (Honig, 2014; Gulrajani and Honig, 2016).

6.5 Public Communications

A move to flexible programming would arguably require a fresh communications approach about the link between financial aid and development (Wild *et al.*, 2015: 44). Many of the challenges associated with adopting more political and flexible programmes in which donor agency staff, evaluators, partners or suppliers are allowed to experiment and learn from good quality monitoring and evaluation data are explained in terms of ‘perverse incentives’ created by the need to be accountable to politicians and donor citizens (Morton, 2009; Camfield *et al.*, 2014; Wild *et al.*, 2015). Therefore, on the face of it, any significant shift to a more cautious, reflexive and relational approach to aid would need high levels of political and public support. Wild *et al.* (2015) argue this would require a fresh communications approach to explaining how aid works and the conditions in which it supports development.

Practitioners generally agree development cooperation agencies have a limited understanding of whether and how different kinds of messaging influence public support and understanding of aid. However, there has been some research to suggest those who support development cooperation would like a more nuanced explanation of what it does (Glennie *et al.*, 2013). There is public fatigue with messages about need (Wild *et al.*, 2015: 44). Despite this, it is recognised there are risks associated with admitting and talking about

⁶⁵ This impression was shared by another researcher who has recently undertaken ethnographic research in a DFID country office and conversations with several ex DFID staff in August 2016.

the contingent conditions in which aid can support real and lasting change (Eyben, 2010). Eyben (2010) argues that explaining the reality of development cooperation to the public could bring the system ‘crashing down’.

As the preceding discussion indicates there is currently more open discussion of potential problems associated with the commitment to allocating 0.7% of gross national income for development cooperation. More money does not necessarily produce better results. Nonetheless differences of opinion and interests between those who have campaigned long and hard for the 0.7% aid effectiveness target and those who see it as a fundamental barrier to doing things differently, for example Booth (2012b), could prove a real obstacle to change. A debate within the sector about the pros and cons of the 0.7% is likely to be bitter and emotional, particularly in the UK where the target is under attack by the media and sections of the general public.⁶⁶ But perhaps public criticism offers opportunities? Even though many practitioners are likely to find DFID’s arguments that aid progresses national interests deplorable,⁶⁷ the links made between aid, global disease and migration provide an entry point into discussions that challenge a transactional view of development. They allow opportunities to talk about the new emphasis on global challenges, like climate change and the roles many donor countries have played in creating or perpetuating them.⁶⁸

6.6 Summary

Shifting to an alternative development paradigm would involve quite substantial change in organisational approaches to policymaking; results based management systems; finance, procurement and

⁶⁶ This story printed by the Daily Mail on 27 March 2016 is one of a series criticising the 0.7% aid target. Daily Mail. (2016). Revealed, how UK aid funds TERRORISTS: After yet more budget cuts, another £12bn of your taxes are being splurged on foreign hand-outs for militants, killers, Palestinian palaces and jobs that don't exist. Daily Mail, [Online], Available: <http://www.dailymail.co.uk/news/article-3510827/Revealed-UK-aid-funds-TERRORISTS-budget-cuts-12bn-taxes-splurged-foreign-hand-outs-militants-killers-Palestinian-palaces-jobs-don-t-exist.html> [3 April 2016].

⁶⁷ DFID. (2016). Media responses to reports on UK aid projects: Setting the record straight. UK Government, [Online], Available: <https://www.gov.uk/government/news/media-reports-on-uk-aid-projects-setting-the-record-straight> [3 April 2016].

⁶⁸ For a discussion about rich countries’ impact on climate change see: Oxfam. (2015). Carbon Inequality: *Why the Paris climate deal must put the poorest, lowest emitting and most vulnerable people first* [Online], Available: https://www.oxfam.org/en/research/extreme-carbon-inequality?utm_source=oxf.am&utm_medium=Zecv&utm_content=redirect

contracting policies; human resource policies and procedures; and public communications. As many of these involve revising long held assumptions about how organisations work, as well as behaviours and norms, they are unlikely to happen quickly and will require considerable debate and challenge.

7. Conclusions and Implications

I started this report with a story about the dilemmas experienced by a MEL Manager trying to decide what kind of indicators to include in a logical framework used for performance management within an adaptive programme committed to learning. A history of the implementation of the results agenda in the UK illuminated the reasons for her concerns, many of which are shared by other practitioners. They worry that a poor choice of indicator, which focuses on meaningless numbers for political accountability, in a complex unpredictable programme could lead to negative impacts on staff and partner morale, or, even worse, suspension of payments. I suggested she consider approaches recommended by groups of practitioners who are piloting and advocating methods that challenge some of the limiting assumptions embedded in the RBM tools she was using.

An analysis of wider arguments made by the various groups criticising short-term RBM approaches indicate they broadly agree that assumptions underpinning the technical tools used to plan, monitor, evaluate and learn in development cooperation relationships are not fit for purpose. Critics of a narrow interpretation of the results agenda concur that donors should move away from best practice models and embrace uncertainty. Efforts to do so have generated useful lessons about the root causes of RBM's weaknesses, as well as the pros and cons of suggested means to overcome them.

7.1 Problematic assumptions

The groups critiquing RBM approaches are driven by varied motivations; however, their arguments expose a number of common flaws in the assumptions underpinning RBM tools. These include the ideas that change is technical, linear and predictable. But they go further and draw attention to the fact that those involved in development relationships often have different theories about the causes of problems, solutions and pathways to change.

Unsurprisingly perhaps, recognition that different actors have different understandings of the aims of development cooperation interventions has led to questions concerning the nature of knowledge, evidence and learning in development cooperation. Who needs to learn what, where and when to achieve, assess and

communicate results for different types of interventions implemented in different political and social contexts? As well as eschewing the notion that policymakers and taxpayers are the sole users of evidence and learning about results, these critics also dispute NPM's assumptions about human behaviour, highlighting the importance of informal relationships, power and social norms. These are forgotten in RBM models that are based on the notion that people are driven by individual interests and financial incentives alone.

7.2 'Alternative' ideas and tools

There is no shortage of ideas about how development cooperation agencies could enhance the way they address the political, complex and interrelated national and global issues facing the world today. A range of planning, monitoring and evaluation methods informed by complex adaptive systems thinking, political economy and power analysis, as well as social theory on relationships, have been developed, some of which are newer than others.

Approaches such as problem driven iterative adaptation (PDIA) are based on assumptions that much development and change is political, unpredictable, non-linear and relies on local leadership and the quality of relationships. Some of this thinking, particularly related to politics and power, has also informed recommendations for the use of theories of social change to make the assumptions of those involved in cooperation relationships more explicit. This is not only to enhance learning about how change happens, but also to encourage more critically reflexive trust-based development cooperation relationships. Other innovations include the development and promotion of monitoring and evaluation approaches that enable flexibility and adaptation, while assessing performance based on outcomes rather than outputs. The importance of contextually situated knowledge and the need to learn during programme implementation is emphasised.

7.3 Challenges putting ideas into practice

Those critiquing dominant approaches to development management have generated useful practical lessons that show it is possible to apply the approaches they have developed to overcome limiting assumptions associated with RBM and achieve better than expected results. But

even though their ideas have gained traction, their experiences illustrate the challenges of trying to institutionalise management approaches that support politically smart, locally led adaptive programmes and transition to an alternative development management paradigm. Efforts to mainstream the use of political economy analysis and theories of change in organisations operating within a new public management paradigm have been disappointing, and much of the work appears to have been gender blind.

There are some good examples of participatory monitoring and evaluation as well as creative approaches to developing results indicators that can be changed in line with programme adaptations that meet donor performance monitoring and programme learning needs. However, agreeing these and making them work often depends on individuals. Scaling up such approaches is difficult because of the increasingly complex and messy power relations between those funding, managing and implementing programmes. The time and costs associated with the relational work of constant negotiations in an era when political pressure means many traditional donors have to appear transparent and manage risk are high. Carefully adapted payment by results type models may occasionally provide such possibilities. But preliminary evidence suggests that concerns about political accountability mean they focus on the wrong results. Thus they are unlikely to provide the panacea to enhancing local ownership and flexibility that some advocates of politically smart adaptive programming expected.

In view of the mixed results described above, demonstrating the efficacy of politically smart, adaptive approaches that enable learning is a priority. But some are concerned that the methods used to generate such 'evidence' could undermine arguments for change if they are embedded in the assumptions underpinning RBM and NPM. Everyone thinks learning is important and more are accepting of pluralist research methodologies. However, discussions of assumptions concerning the nature and purposes of evidence and learning for different actors are still quite nascent. This may need addressing to avoid the risk of ideas from DDD, TWP and other groups becoming seen as well-intentioned top-down initiatives based on new wine in old bottles.

7.4 The need for institutional and organisational change

Moving to the institutional change required to support an alternative management paradigm, considerable challenges lie ahead. New performance monitoring, evaluation and contracting tools are not sufficient. Those critiquing RBM argue that aid agencies need to be viewed as contingent systems, rather than bureaucratic machines as is assumed by NPM theory before a number of issues can be addressed. These include: (1) revising communications strategies and messages for politicians and the general public; (2) redesigning performance management and monitoring evaluation and learning systems; (3) additional research about the effects of leadership and human resource policy on autonomy, trust and development or social change outcomes; and (4) re-conceptualising cost relationships, contracting procedures and value for money frameworks.

Experiences from DFID's attempts to make progress in a couple of the above areas illustrate how difficult it is to shift organisational culture and behaviour. A politically informed complexity lens suggests change cannot be orchestrated at a systemic level by policy change and minor alterations to results based management tools and incentives alone. The limitations of principal agent theory need to be confronted. Changing norms and behaviours in development agencies involves challenging much more diffuse forms of power and the informal rules of the game. Policymakers and practitioners who are impatient for a rapid shift from a top-down results based development management culture, to one that is truly locally led, are likely to be disappointed in the short-term.

There is considerable support for alternative development management approaches and many people are already working with politically and complexity informed lenses. Yet because influencing wider change in development management involves re-evaluating the political ideas and assumptions underpinning many aid agency policy and management systems it is likely to involve heated exchange. According to some proponents of complexity thinking, like Mowles (2011) these are the sites of struggle that really matter. Change is the unpredictable consequence of social interaction and relationships. Hence focusing and reflecting on what is going on in everyday practice and conversations about means to measure and enhance results is as important as making grand plans for institutional reform. Those involved could do worse than take advice from Mowles (2011)

and participants attending the Politics of Evidence and Results conference hosted by the Big Push Forward to focus on considering how power relations in their various local interactions influence the take-up of ideas in programmes or organisational change.⁶⁹

7.5 Implications

Policymakers need to recognise the limitations of traditional RBM tools underpinned by unrealistic assumptions that obscure the complex political challenges as experienced by those involved in development cooperation relationships at country and global levels. However, there are no magic bullets when it comes to finding management approaches that will enable the achievement and measurement of rights based social change. Thus policymakers need to interrogate assumptions underpinning approaches promoted as offering new solutions that may have been oversold, such as payment for results models.

Since complexity informed adaptive learning models may struggle to meet the expectations of management systems constrained by NPM's bureaucratic norms, policymakers can play useful roles in supporting policy research and recommendations informed by contingency theory. These highlight the potential benefits of testing performance management methods that value trust, staff autonomy and learning over disempowering techniques for the precise measurement of results.

As well as supporting ongoing efforts to identify and demonstrate the effectiveness of alternative theoretical and practical approaches to programme and aid agency management, policymakers would be well advised to support work related to developing new public communication strategies. Critics often argue that the RBM tools that donors use to collect data on results that can be communicated to the public may undermine domestic accountability rather than enhance it. Embedded in a logic that implies social change is predictable and can be managed, they obscure the complex real world challenges as experienced by those involved in development cooperation relationships at country and global levels.

⁶⁹ Details of the conference can be found in Eyben, R., Guijt, I., Roche, C., Shutt, C. and Whitty, B. (2014). Politics of Evidence Conference Report. Big Push Forward, [Online], Available: <http://bigpushforward.net/wp-content/uploads/2013/09/BPF-PoE-conference-report.pdf> [Accessed 4 April 2016].

If a shift in paradigm is to be realised policymakers need to promote complexity informed public communications strategies that paint a more nuanced picture of what is at stake. This will involve supporting research that explores the potential for encouraging the public and politicians to think about results and risks in relation to interrelated global issues like migration and climate change and the potential advantages of long-term solidarity approaches over short-term results to tackle them. Pursuing such arguments is likely to be challenging. But having debates about difficult issues, such as whether to frame arguments in ways that appeal to national interests or more progressive social norms and what that might mean about the wisdom of the 0.7% aid targets is, according to some complexity theorists, a necessary part of a global systems change process.

References

- ActionAid International. (2012). *People's action to end poverty*. UK: ActionAid, [Online], Available: http://www.actionaid.org/sites/files/actionaid/peoples_action_to_end_poverty_28pp.pdf [3 April 2016]
- Abu Alghaib, O. (2015). "Aid bureaucracy and support for disabled people's organisations: a fairy tale of self-determination and self-advocacy". In (eds) Eyben, R., Guijt, I., Roche, C. and Shutt, C., *The Politics of Evidence in International Development: Playing the Game to Change the Rules*, pp.115-134. Rugby, UK: Practical Action.
- Andrews, M., Pritchett, L. and Woolcock, M. (2012). *Escaping capability traps through problem driven iterative adaptation (PDIA)*. WIDER Working Paper 2012/064. Helsinki: UNU-WIDER.
- Argyris, C. and Schon, D. A. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco: Jossey-Bass.
- Bamberger, M., Tarsilla, M. and Hesse-Biber, S. (2016). Why so many "rigorous" evaluations fail to identify unintended consequences of development programs: how mixed methods can contribute. *Evaluation and Program Planning*, 55, pp.155-162.
- Bastoe, P. O. (2006). "Implementing results-based management". In (eds) Rist, R. C. and Stame, N., *From studies to streams: Managing evaluative systems* (Vol. 12), pp.97-112. Piscataway, USA: Transaction Publishers.
- Bateson, G. (1972). *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. Chicago, USA: University of Chicago Press.
- Binnendijk, A. (2001). *Result-based management in the development co-operation agencies: A review of experience: Background report*. Paris, France: Organisation for Economic Co-operation and Development.
- Booth, D. (2012a). *Synthesis report: Development as a collective action problem: Addressing the real challenges of African governance*. Africa Power and Politics Programme. London, UK: Overseas Development Institute.

- Booth, D. (2012b). Aid effectiveness: bringing country ownership (and politics) back in. *Conflict, Security & Development*, 12(5), pp.537-558.
- Booth, D. (2013). *Facilitating development: An arm's length approach to aid*. London, UK: Overseas Development Institute.
- Booth, D. and Chambers, V. (2014). *The SAVI programme in Nigeria*. Discussion Paper. London, UK: Overseas Development Institute.
- Booth, D. and Unsworth, S. (2014). *Politically smart, locally led development*. Discussion Paper. London, UK: Overseas Development Institute.
- Brolin, T. (Forthcoming). Framing the results agenda in Swedish development co-operation. *Development Policy Review*.
- Buffardi, A. L. (2016). *When theory meets reality: Assumptions, feasibility and implications of a complexity-informed approach*. A Methods Lab publication. London, UK: Overseas Development Institute.
- Camfield, L., Duvendack, M. and Palmer-Jones, R. (2014). Things you wanted to know about bias in evaluations but never dared to think. *IDS Bulletin*, 45(6), pp.49-64.
- Carothers, T. and Gramont, D. de. (2013). *Development aid confronts politics: The almost revolution*. New York: Carnegie Endowment for International Peace.
- Causemann, B. and Gohl, E. (2015). "Unwritten reports: Lessons from an NGO collective". In (eds) Eyben, R., Guijt, I., Roche, C. and Shutt, C. *The Politics of Evidence in International Development: Playing the Game to Change the Rules*, pp.173-192. Rugby, UK: Practical Action.
- Chambers, R. (2010). *Paradigms, poverty and adaptive pluralism*. IDS Working Paper 334. Brighton, UK: Institute of Development Studies.
- Chambers, R. (2015). Inclusive rigour for complexity. *Journal of Development Effectiveness*, 7(3), pp.327-335.
- Cracknell, B. (1996). Evaluating development aid, *Evaluation*, 2(1) 23-33.

- Dan, S. and Pollitt, C. (2015). NPM can work: An optimistic review of the impact of New Public Management reforms in central and eastern Europe. *Public Management Review*, 17(9), pp.1305-1332.
- Dar, S. and Cooke, B. (2008). *The new development management: Critiquing the dual modernization*. UK: Zed Books.
- DFID. (2015). *Department for International Development annual report and accounts 2014–2015*. UK: DFID, [Online], Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/445698/Annual-Report-2014-2015.pdf [31 March 2016].
- DFID and The UK Treasury. (2016). *UK Aid: Tackling global challenges in the national interest*. UK Government, [Online], Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/478834/ODA_strategy_final_web_0905.pdf [31 March 2016].
- DuBois, M., Wake, C., Sturridge, S. and Bennett, C. (2015). *The Ebola response in West Africa: Exposing the politics and culture of international aid*. HPG Working Paper. London, UK: Overseas Development Institute.
- Earl, S., Carden, F. and Smutylo, T. (2001). *Outcome mapping: Building learning and reflection into development programs*. Ottawa, Canada: International Development Research Center.
- Ebrahim, A. (2003). Accountability in practice: Mechanisms for NGOs. *World Development*, 31(5), pp.813-829.
- Ebrahim, A. (2005). *NGOs and organizational change: Discourse, reporting, and learning*. Cambridge, UK: Cambridge University Press.
- Es, M. van, and Guijt, I. (2015). “Theory of change as best practice or next trick to perform? Hivos’ journey with strategic reflection”. In (eds) Eyben, R., Roche, C., Guijt, I. and Shutt, C., *The Politics of Evidence in International Development: Playing the Game to Change the Rules*, pp.95-114. Rugby, UK: Practical Action Ltd.
- Eyben, R. (2003). *Donors as political actors: Fighting the Thirty Years War in Bolivia*. Institute of Development Studies Working Paper 183. Brighton, UK: Institute of Development Studies.

- Eyben, R. (2005). Donors' learning difficulties: Results, relationships and responsibilities. *IDS Bulletin*, 36(3), pp.98-107.
- Eyben, R. (2006a). The power of the gift and the new aid modalities. *IDS Bulletin*, 37(6). pp.88-98.
- Eyben, R. (Ed.). (2006b). *Relationships for aid*. UK and USA: Earthscan.
- Eyben, R. (2008). *Power, mutual accountability and responsibility in the practice of international aid: A relational approach*. IDS Working Paper 305. Brighton, UK: Institute of Development Studies.
- Eyben, R., Kidder, T., Rowlands, J. and Bronstein, A. (2008). Thinking about change for development practice: A case study from Oxfam UK. *Development in Practice*, 18(2), pp.201-212.
- Eyben, R. (2010). Hiding relations: The irony of "effective aid". *European Journal of Development Research*, 22(3), pp.382-397.
- Eyben, R. (2013). *Uncovering the politics of 'evidence' and 'results'. A framing paper for development practitioners*. Brighton, UK: Institute of Development Studies.
- Eyben, R., Guijt, I., Roche, C., Shutt, C. and Whitty, B. (2014). *Politics of Evidence Conference Report*. Big Push Forward, [Online], Available: <http://bigpushforward.net/wp-content/uploads/2013/09/BPF-PoE-conference-report.pdf> [Accessed 4 April 2016].
- Eyben, R. and Guijt, I. (2015). "Introduction". In (eds) Eyben, R., Guijt, I., Roche, C. and Shutt, C. *The Politics of Evidence in International Development: Playing the Game to Change the Rules*. pp.1-18. Rugby, UK: Practical Action.
- Eyben, R. (2015). "Uncovering the politics of evidence and results". In (eds) Eyben, R., Guijt, I., Roche, C. and Shutt, C. *The Politics of Evidence in International Development: Playing the Game to Change the Rules*, pp.19-38. Rugby, UK: Practical Action.
- Fisher, J. and Marquette, H. (2014). *Donors doing political economy analysis: From process to product (and back again?)*. Working Paper, Developmental Leadership Program. Birmingham: University of Birmingham.

- Fisher, J. and Marquette, H. (2016). “Empowered patient” or “doctor knows best”? Political economy analysis and ownership. *Development in Practice*, 26(1), pp.115-126.
- Forss, K., Marra, M. and Schwartz, R. (Eds). (2011). *Evaluating the complex: Attribution, contribution, and beyond* (Vol. 18). Piscataway, USA: Transaction Publishers.
- Furubo, J. E., Rist, R. C. and Speer, S. (Eds). (2013). *Evaluation and turbulent times: Reflections on a discipline in disarray* (Vol. 20). Piscataway, USA: Transaction Publishers.
- Gold, J., Wilson-Grau, R. and Fisher, S. (2014). *Cases in outcome harvesting: Ten pilot experiences identify new learning from multi-stakeholder projects to improve results*, Washington, DC: World Bank Group.
- Gonsalvez De Asis, M. and Woolcock, M. (2015). *Operationalizing the science of delivery agenda to enhance development results*. Washington, DC: The World Bank.
- Global Partners Governance. (2014). *The logframe and the beautiful game: Project logic v football logic*. London, UK: Global Partners Governance.
- Glennie, J., McKechnie, A., Rabinowitz, G. and Ali, A. (2013). *Localising aid: Sustaining change in the public, private and civil society sectors*. Centre for Aid and Public Expenditure. London, UK: Overseas Development Institute.
- Green, D. (Forthcoming). *How change happens*. Oxford, UK: Oxford University Press.
- Guijt, I. (2008). *Seeking surprise: Rethinking monitoring for collective learning in rural resource management*. Unpublished PhD dissertation. Wageningen University.
- Guijt, I. (2015). “Playing the rules of the game and other strategies”. In (eds) Eyben, R., Guijt, I., Roche, C. and Shutt, C. *The Politics of Evidence in International Development: Playing the Game to Change the Rules*, pp.193-210. Rugby, UK: Practical Action.
- Gulrajani, N. (2015). Dilemmas in donor design: Organisational reform and the future of foreign aid agencies. *Public Administration and Development*, 35(2), pp.152-164.

- Gulrajani, N., & Honig, D. (2016). *Reforming donors in fragile states*. Report. London, UK: Overseas Development Institute.
- Haynes, P. (2015). *Managing complexity in the public services*. UK: Routledge.
- Honig, D. (2014). *Navigation by judgment: Organizational autonomy and country context in the delivery of foreign aid*. Kennedy School of Government, Harvard University: Cambridge, MA.
- Holland, J. (Ed.). (2013). *Who counts? The power of participatory statistics*. Rugby, UK: Practical Action Publications Ltd.
- Holzapfel, S. (2016). Boosting or hindering aid effectiveness? An assessment of systems for measuring donor agency results. *Public Administration and Development*, 36(1), pp.3-19.
- Hudson, D. and Marquette, H. (2015). “Mind the gaps: What’s missing in political economy analysis and why it matters”. In (eds) Whaites *et al.*, *A Governance Practitioner’s Notebook: Alternative Ideas and Approaches*, OECD.
- Hudson, D., Marquette, H. and Waldo, S. (2016). *Everyday Political Economy Analysis*. Developmental Leadership Program, University of Birmingham. Birmingham: DLP.
- Hughes, K. (2012). NGOs and the results agenda: Unrealistic expectations or a blessing in disguise? *Norrag News*, 47, pp.80–82.
- Hummelbrunner, R. (2015). Learning, Systems Concepts and Values in Evaluation: Proposal for an Exploratory Framework to Improve Coherence. *IDS Bulletin*, 46(1), pp.17-29.
- Hummelbrunner, R. and Jones, H. (2013). *A Guide for Managing in the Face of Complexity*. Working Paper. London, UK: Overseas Development Institute.
- Hutchings, C. (2014). Balancing accountability and learning: A review of Oxfam GB’s global performance framework. *Journal of Development Effectiveness*, 6(4), pp.425-435.
- Independent Commission for Aid Impact. (2014). How DFID learns, UK: ICAI. [Online], Available: <http://icai.independent.gov.uk/wp-content/uploads/How-DFID-Learns-FINAL.pdf> [3 April 2016].
- Independent Commission for Aid Impact. (2015). How DFID delivers impact, UK: ICAI. [Online], Available:

<http://icai.independent.gov.uk/wp-content/uploads/ICAI-report-DFIDs-approach-to-Delivering-Impact.pdf> [3 April 2016].

- IDL Group GRM. (2014). Monitoring and learning in politically smart and adaptive programmes. Workshop note 11 December 2014. Applied Political Economy Analysis Group. London, UK: IDL Group GRM.
- Jabeen, S. (2016). Do we really care about unintended outcomes? An analysis of evaluation theory and practice. *Evaluation and Program Planning*, 55, pp.144-154.
- Jupp, D., Sohel Ibn, A. and Barahona, C. (2010). *Measuring Empowerment? Ask Them: Quantifying Qualitative Outcomes from People's Own Analysis: Insights for Results-based Management from the Experience of a Social Movement in Bangladesh*. Stockholm, Sweden: Sida.
- Kleinfeld, R. (2015). *Improving development aid design and evaluation: Plan for sailboats, not trains*. Washington, DC: Carnegie Endowment for International Peace.
- Krznicar, R. (2007). *How change happens: Interdisciplinary perspectives for human development*. Oxfam Research Report. Oxford, UK: Oxfam.
- Kusek, J. Z. and Rist, R. C. (2004). *Ten steps to a results-based monitoring and evaluation system: A handbook for development practitioners*. World Bank Publications.
- Ladner, D. (2015). *Strategy testing: An innovative approach to monitoring highly flexible aid programs*. San Francisco, CA: The Asia Foundation.
- Leroy, M. (2012). The aid industry is threatening partner countries with its results obsession disorder. *Norrag News*, 47, pp.55–6.
- Lipskey, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York, USA: Russel Sage Foundation.
- Lloyd, R., Poate, D. and Villanger, E. (2014). Results measurement and evaluability: A comparative analysis. *Journal of Development Effectiveness*, 6(4), pp.378-391.

- Maclay, C. (2015). Management not models: Adaptability, responsiveness, and a few lessons from football. *Development in Practice*, 25(1), pp.42-57.
- Manning, R. and White, H. (2014). Measuring results in development: The role of impact evaluation in agency-wide performance measurement systems. *Journal of Development Effectiveness*, 6(4), pp.337-349.
- Management for Development Results. (2004). MfDR principles in action: Sourcebook on emerging good practices, [Online], Available: <http://www.mfdr.org/Sourcebook/1stEdition/4-MfDRPrinciples.pdf> [4 April 2016].
- Mayne, J. (2007). Challenges and lessons in implementing results-based management. *Evaluation*, 13(1), pp.87-109.
- Mayne, J. and Rist, R. C. (2006). Studies are not enough: The necessary transformation of evaluation. *The Canadian Journal of Program Evaluation*, 21(3), pp.93-120.
- Mercy Corps. (2015). *Managing complexity: Adaptive management at Mercy Corps*. Mercy Corps, [Online], Available: https://d2zyf8ayvg1369.cloudfront.net/sites/default/files/Adaptive%20management%20paper_external.pdf [4 April 2016].
- Mowles, C., Stacey, R. and Griffin, D. (2008). What contribution can insights from the complexity sciences make to the theory and practice of development management? *Journal of International Development*, 20(6), pp.804-820.
- Mowles, C. (2011). *Rethinking management: Radical insights from the complexity sciences*. Aldershot, UK: Gower Publishing Ltd.
- Mowles, C. (2014). Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship. *Evaluation*, 20(2), pp.160-175.
- Morton, J. (2009) *Why We Will Never Learn: A political economy of aid effectiveness*. London: James Morton & Co.
- Mosse, D., Farrington, J. and Rew, A. (1998). *Development as process: Concepts and methods for working with complexity*. London, UK: Routledge and Overseas Development Institute.

- Natsios, A. (2010). *The clash of the counter-bureaucracy and development*. Center for Global Development Essay. Washington DC, USA: Center for Global Development.
- Nielsen, S. B. and Ejler, N. (2008). Improving performance? Exploring the complementarities between evaluation and performance management. *Evaluation*, 14(2), pp.171-192.
- O’Keefe, M., Sidel, J. T., Marquette, H., Roche, C., Hudson, D. and Dasandi, N. (2014). *Using action research and learning for politically informed programming*, DLP Research Paper 29. University of Birmingham, UK: Developmental Leadership Program.
- O’Neil, T. (2016). *Using adaptive development to support feminist action*. Report. London, UK: Overseas Development Institute.
- Organisation for Economic Co-operation and Development. (2010). *Glossary of key terms in evaluation and results based management*. OECD, [Online], Available: <http://www.oecd.org/development/peer-reviews/2754804.pdf> [4 April 2016].
- Organisation for Economic Co-operation and Development. (2013). *Development cooperation peer reviews: Sweden report*. OECD, [Online], Available: <http://www.oecd.org/dac/peer-reviews/sweden-peer-review-2013.pdf> [6 April 2016].
- Organisation for Economic Co-operation and Development. (2014). *Development cooperation peer reviews: United Kingdom report*. OECD, [Online], Available: <http://www.oecd.org/dac/peer-reviews/UK%20peer%20review%202014.pdf> [4 April 2016].
- Organisation for Economic Co-operation and Development. (2015). *Development cooperation peer reviews: Germany report*. OECD, [Online], Available: <http://www.oecd.org/dac/oecd-development-co-operation-peer-reviews-germany-2015-9789264246133-en.htm> [4 April 2016].
- Oxfam. (2015). *Carbon Inequality: Why the Paris climate deal must put the poorest, lowest emitting and most vulnerable people first*. Oxford: UK: Oxfam, [Online], Available: https://www.oxfam.org/en/research/extreme-carbon-inequality?utm_source=oxf.am&utm_medium=Zecv&utm_content=redirect

- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York City, USA: Guilford Press.
- Picciotto, R. (2016). Review of *The Politics of Evidence and Results in International Development: Playing the Game to Change the Rules?* by Rosalind Eyben, Irene Guijt, Chris Roche, and Cathy Shutt. Rugby, UK: Practical Action Publishing, 2015. *Governance*, 29, pp.147–149.
- Pinnington, R. (2014). *Local First in Practice; Unlocking the power to get things done*. London: Peace Direct.
- Pollitt, C. and Dan, S. (2013). Searching for impacts in performance-oriented management reform: A review of the European literature. *Public Performance & Management Review*, 37(1), pp.7-32.
- Power, G. and Coleman, O. (2011). *The Challenges of Political Programming*. Discussion Paper. Stockholm, Sweden: International IDEA.
- Pritchett, L., Samji, S. and Hammer, J. S. (2013). *It's all about MeE: Using structured experiential learning ('e') to crawl the design space*. Working Paper 322. Washington DC, USA: Center for Global Development.
- Ramalingham, B. (2013). *Aid on the Edge of Chaos*. Oxford, UK: Oxford University Press.
- Ramalingham, B., Laric, M. and Primrose, J. (2014). *From best practice to best fit: Understanding and navigating wicked problems in international development*. Working Paper. London, UK: Overseas Development Institute.
- Renzio, P. de, Booth, D., Rogerson, A. and Curran, Z. (2005). *Incentives for harmonisation and alignment in aid agencies*. London, UK: Overseas Development Institute.
- Reynolds, M. (2015). (Breaking) The iron triangle of evaluation. *IDS Bulletin*, 46(1), pp.71-86.
- Rist, R. C., and Stame, N. (eds). (2006). *From Studies to Streams: Managing Evaluative Systems*, (Vol. 12), Piscataway, USA: Transaction Publishers.
- Roche, C. (2015). "The politics of juggling multiple accountability disorder". In (eds) Eyben, R., Roche, C., Guijt, I. and Shutt, C.,

- The Politics of Evidence in International Development: Playing the Game to Change the Rules*, pp.79-94. Rugby, UK: Practical Action.
- Rogers, P. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, 14(1), pp.29-48.
- Root, H., Jones, H. and Wild, L. (2015) *Managing Complexity and Uncertainty in Development Policy and Practice*. Report. London, UK: Overseas Development Institute.
- SAVI. (2014). *Rising to the challenge: Supporting “problem driven iterative adaptation” and “politically smart, locally led” approaches through a donor-funded programme: The experience of the State Accountability and Voice Initiative in Nigeria*. SAVI, [Online], Available: http://savi-nigeria.org/wp-content/uploads/2014/10/DFID_SAVI_brief_Challenge_Online.pdf [3 April 2016].
- SAVI. (2015). *Defining and measuring results: monitoring incremental change over time*. Approach Paper. SAVI, [Online], Available: http://savi-nigeria.org/wp-content/uploads/2015/12/SAVI_ApproachPaper5_2015_FINAL.pdf [3 April 2016].
- Shah, R. (2014). Assessing the “true impact” of development assistance in the Gaza Strip and Tokelau: “Most Significant Change” as an evaluation technique. *Asia Pacific Viewpoint*, 55(3), pp.262–276.
- Shutt, C. (2006). “Money matters in aid relationships.” In (ed) Eyben.R., *Relationships for aid*. UK and USA: Earthscan.
- Shutt, C. and McGee, R. (2013). *Improving the evaluability of INGO empowerment and accountability programmes*. Centre for Development Impact Practice Paper 1. Brighton, UK: Institute of Development Studies.
- Shutt, C. (2015). “The Politics and Practice of Value for Money”. In (eds) Eyben, R., Roche, C., Guijt, I. and Shutt, C., *The Politics of Evidence in International Development: Playing the Game to Change the Rules*, pp.57-78. Rugby, UK: Practical Action.
- Sida. (2015). *Results based financing approaches: What are they? A methodological introduction*. Stockholm, Sweden: Sida.

- Sjöstedt, M. (2013). Aid effectiveness and the Paris Declaration: A mismatch between ownership and results based management? *Public Administration and Development*, 33(2), pp.143-155.
- Statskontoret. (2013). *Aid through Swedish organisations: A review of the framework agreement system*. Stockholm, Sweden: Swedish Agency for Public Management.
- Stein, D. and Valters, C. (2012). *Understanding “theory of change” in international development: A review of existing knowledge*. The Asian Institute and the Justice and Security Research Programme.
- Stern, E. (2008). *The Paris Declaration, aid effectiveness and development effectiveness*. The Ministry of Foreign Affairs, Denmark.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. and Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations*. Report of a study commissioned by DFID. UK: DFID.
- Townsend, J. G., Porter, G. and Mawdsley, E. (2002). The role of the transnational community of non-government organizations: governance or poverty reduction? *Journal of International Development*, 14(6), pp.829-839.
- United Nations Development Programme. (2013). *Innovations in monitoring & evaluating results* Discussion Paper. New York, USA: United Nations Development Programme.
- Vähämäki, J., Schmidt, M. and Molander, J. (2011). *Review: Results based management in development cooperation*. Stockholm, Sweden: Riksbankens Jubileumsfond.
- Vähämäki, J. (2015). “The results agenda in Swedish development cooperation: Cycles of failure or reform success?” In (eds) Eyben, R., Roche, C., Guijt, I. and Shutt, C., *The Politics of Evidence in International Development: Playing the Game to Change the Rules*, pp.135-154. Rugby, UK: Practical Action Ltd.
- Vallejo, B., & Wehn, U. (2016). Capacity development evaluation: The challenge of the results agenda and measuring return on investment in the Global South. *World Development*, 79, pp.1-13.

- Valters, C. (2015). *Theories of change: Time for a radical approach to learning in development*. London: Overseas Development Institute and The Asia Foundation.
- Valters, C., Cummings, C. and Nixon, H. (2016). *Putting learning at the centre: Adaptive development programming in practice*. Report. London, UK: Overseas Development Institute.
- Van Hemelrijck, A., & Guijt, I. (2016). *Balancing Inclusiveness, Rigour and Feasibility: Insights from Participatory Impact Evaluations in Ghana and Vietnam*. CDI Practice Paper 14, Brighton, UK: Institute of Development Studies.
- Van Ongevalle, J., Huyse, H., and Van Petegem, P. (2014). Dealing with complexity through actor-focused planning, monitoring and evaluation (PME). *Evaluation*, 20(4), pp. 447-466
- Wallace, T., Bornstein, L., and Chapman, J. (2006). *The aid chain: Commitment and coercion in development NGOs*. Rugby: ITDG Publishing.
- Wallace, T. and Porter, F. (2013). "Introduction". In (eds) Wallace, T., Porter, F. and Ralph-Bowman, M., *Aid, NGOs and the Realities of Women's Lives: A Perfect Storm*. Rugby, UK: Practical Action Publishing Ltd.
- Welle, K. (2013). *Monitoring performance or performing monitoring: The case of rural water access in Ethiopia*. Unpublished PhD dissertation, University of Sussex.
- Westhorp, G. (2014). *Realist impact evaluation: An introduction*. Methods Lab. London: Overseas Development Institute.
- Whitty, B. S. (2015). "Mapping the results landscape: Insights from a crowdsourcing survey". In (eds) Eyben, R., Roche, C., Guijt, I. and Shutt, C., *The Politics of Evidence in International Development: Playing the Game to Change the Rules*, pp.57-78. Rugby, UK: Practical Action Ltd.
- Win, E. (2004). "If it doesn't fit on the blue square it's out!" An open letter to my donor friend". In (eds) Groves, L. and Hinton, R., *Inclusive Aid: Changing Power and Relationships in International Development*, pp.123-127. Sterling VA: Earthscan.

- Woolcock, M. (2013). Using case studies to explore the external validity of ‘complex’ development interventions. *Evaluation*, 19(3), pp.229-248.
- World Bank Group. (2015). “Adaptive designs, adaptive interventions”. In *World Development Report 2015: Mind, Society, and Behavior*. Chapter 11, pp.192-201. Washington, DC: World Bank.
- Wild, L., Booth, D., Cummings, C., Foresti, M. and Wales, J. (2015). *Adapting development: Improving services to the poor*. Report. London: Overseas Development Institute.
- Yanguas, P. and Hulme, D. (2015). Barriers to political analysis in aid bureaucracies: From principle to practice in DFID and the World Bank. *World Development*, 74, pp.209-219.

Website Resources

- Aid Leap. (2016). I analysed 600 of DFID’s annual review and here’s what I found, [Online], Available: <http://aidleap.org/2016/03/14/i-analysed-600-of-dfids-annual-reviews-heres-what-i-found/> [20 March 2016].
- Barder, O. (2012a). What are the results agenda? Owen abroad, [Online], Available: <http://www.owen.org/blog/5228> [4 April 2016].
- Barder, O. (2012b). Seven worries about focusing on results, and how to manage them. Owen abroad, [Online], Available: <http://www.owen.org/blog/5483> [4 April 2016].
- Barder, O. (2014). New research shows aid agencies get better results if they stop trying to control their people on the ground, especially in complex environments (and performance monitoring makes it worse). From poverty to power, [Online], Available: <https://oxfamblogs.org/fp2p/new-research-shows-aid-agencies-get-better-results-if-they-stop-trying-to-control-their-people-on-the-ground-especially-in-complex-environments-and-performance-monitoring-can-make-it-worse/> [3 April 2016].
- Barder, O., Perakis, R., Savedoff, W. and Talbot, T. (2014). 12 principles for payment by results (PbR) in the real world. Center for Global Development, Views from the Center, [Online],

- Available: <http://www.cgdev.org/blog/12-principles-payment-results-pbr-real-world-0> [4 April 2016].
- Barr, J. (2015). Monitoring and evaluating flexible and adaptive programming. Itad, [Online], Available: <http://www.itad.com/monitoring-and-evaluating-flexible-and-adaptive-programming/> [3 April 2016].
- Better Evaluation (undated) Outcome Harvesting. [Outline], Available: http://betterevaluation.org/plan/approach/outcome_harvesting. [4 April 2016].
- Daily Mail. (2016). Revealed, how UK aid funds TERRORISTS: After yet more budget cuts, another £12bn of your taxes are being splurged on foreign hand-outs for militants, killers, Palestinian palaces and jobs that don't exist. Daily Mail, [Online], Available: <http://www.dailymail.co.uk/news/article-3510827/Revealed-UK-aid-funds-TERRORISTS-budget-cuts-12bn-taxes-splurged-foreign-hand-outs-militants-killers-Palestinian-palaces-jobs-don-t-exist.html> [3 April 2016].
- DFID Development Tracker. [Online], Available: <https://devtracker.dfid.gov.uk> [3 April 2016].
- DFID. (2016). Media responses to reports on UK aid projects: Setting the record straight. UK Government, [Online], Available: <https://www.gov.uk/government/news/media-reports-on-uk-aid-projects-setting-the-record-straight> [3 April 2016].
- DFID. (2015). Annual Review Report Sheet for Results-Based Financing for Low Carbon Energy. DFID, [Online], available: bit.ly/1T6nRuG [3 April 2016].
- DFID. (2016). SMART rules: better programme delivery. UK Government, [Online], Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/472771/Smart_Rules_1November_to_31_March_2016.pdf [3 April 2016].
- Doing Development Differently. [Online], Doing Development Differently: The DDD Manifesto Community, Available: <http://doingdevelopmentdifferently.com> [3 April 2016].

- Global Integrity. (2016). Trying, learning and adapting at the GPSA global partners forum, [Online], Available: <http://www.globalintegrity.org/2016/05/trying-learning-adapting-gpsa-global-partners-forum/> [20 May 2016].
- Green, D. (2012). Theories of change = logframes on steroids? A discussion with DFID. From poverty to power. <https://oxfamblogs.org/fp2p/theories-of-change-logframes-on-steroids-a-discussion-with-dfid/> [3 April 2016].
- Green, D. (2015) 'How can research help promote empowerment and accountability?' From Poverty to power, [Online], Available: <http://oxfamblogs.org/fp2p/how-can-research-help-us-promote-empowerment-and-accountability/> [4 April 2016].
- Green, D. (2016). Payment by results hasn't produced much in the way of results but aid donors are doing it anyway – why? From poverty to power, [Online], Available: <http://oxfamblogs.org/fp2p/payment-by-results-hasnt-produced-much-in-the-way-of-results-but-aid-donors-are-doing-it-anyway-why/> [4 April 2016].
- Itad. (2015). Empowerment and accountability macro evaluation, [Online], Available: <http://www.itad.com/knowledge-and-resources/dfids-macro-evaluations/empowerment-and-accountability-macro-evaluation/> [4 April 2016].
- Lowcock, M. and Blackall, D. (2015). Letter of Appointment, UK Government, [Online], Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/415697/Letter-of-appointment-SRO-StHelena-airport.pdf [3 April 2016].
- Maxwell, S. (2014). Is teacher pleased? The DAC peer review of UK 2014, [Online], Available: <http://www.simonmaxwell.eu/blog/is-teacher-pleased-the-dac-peer-review-of-the-uk-2014.html> [3 April 2016].
- Moses M. (2016). Trying, learning, adapting at the GPSA Global Partners forum. Global Integrity: Data, learning and action in open governance, [Online], Available: <http://www.globalintegrity.org/2016/05/trying-learning-adapting-gpsa-global-partners-forum/> [20 May 2016].

- Patton, M.Q. (2010). Developmental evaluation. Better Evaluation, [Online], Available: http://betterevaluation.org/plan/approach/developmental_evaluation [3 April 2016].
- Patton, M.Q. (2016). Blue Marble Evaluation, [Online], Available: www.utilization-focusedevaluation.org/blue-marble-evaluators/ [3 April 2016].
- PowerCube. (2010). [Online], Available: <http://www.powercube.net> [3 April 2016].
- Positive Deviance. [Online], Available: <http://www.positivedeviance.org> [3 April 2016].
- Roche, C. (2012). Results, value for money and the aid budget, [Online], Available: <http://devpolicy.org/results-value-for-money-and-the-aid-budget20120516/> [4 April 2016].
- Shah, R. and Gerson, M. (2016). What do we mean by results based development? Centre for Global Development, [Online], Available: <http://www.cgdev.org/blog/what-do-we-mean-results-based-development-podcast-raj-shah-and-michael-gerson> [31 March 2016].
- Shutt, C. (2011). The value for money conundrum, Big Push Forward, [Online], Available: <http://bigpushforward.net/archives/772> [6 April 2016].
- Snowdon, D. (2010). Cynefin framework, [Online], Available: <https://www.youtube.com/watch?v=N7oz366X0-8> [3 April 2016].
- The Big Push Forward, [Online], Available: website: <http://bigpushforward.net> [3 April 2016].
- Thinking and Working Politically Community of Practice, DLP, [Online], Available: <http://www.dlprog.org/research/thinking-and-working-politically-community-of-practice.php> [3 April 2016].
- United Nations Development Programme. (1997). Who are the question-makers? A participatory evaluation handbook. Office of Evaluation and Strategic Planning, UNDP. [Online] Available <http://web.undp.org/evaluation/documents/wintro.htm> [3 April 2016].

- Watts, C. (2016). What role for evidence in the new aid strategy? Institute of Development Studies, [Online], Available: http://www.ids.ac.uk/events/research-and-knowledge-at-dfid-what-role-for-evidence-in-the-new-aid-strategy?utm_content=bufferd49f7&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer [31 March 2016].
- Wingfield, T. and Vowles, P. (2014). DFID is changing its approach to better address the underlying causes of poverty and conflict – can it work? Guest Post from two DFID reformers. From poverty to power, [Online], Available: <http://oxfamblogs.org/fp2p/dfid-is-changing-its-approach-to-better-address-the-underlying-causes-of-poverty-and-conflict-can-it-work-guest-post-from-tom-wingfield-and-pete-vowles/> [3 April 2016].

Informal Interviews

Elbereth Donavan	LASER Programme
Jonathan Francis	Sida
Rebecka Kitzing-Ivarsson	Sida
Craig Valters	ODI
Pete Vowles	DFID

Abbreviations

ADB	African Development Bank
BPF	Big Push Forward
DAC	Development Assistance Committee
DDD	Doing Development Differently
DFAT	Department of Foreign Affairs and Trade
DFID	Department for International Development
DLP	Development Leadership Programme
ICAI	Independent Commission for Aid Impact
IDEA	Institute for Democracy and Electoral Assistance
IFAD	International Fund for Agricultural Development
INGO	International non-governmental organisation
MDGs	Millennium Development Goals
MEL	Monitoring evaluation and learning
MfDR	Managing for development results
MSC	Most Significant Change
NGO	Non-governmental organisation
NPM	New public management
ODI	Overseas Development Institute
OECD	Organisation for Economic Development
PBR	Payment by results
PDIA	Problem Driven Iterative Adaptation
PEA	Political economy analysis
PIALA	Participatory Impact Assessment and Learning Approach
PM	Participatory methodology
RBM	Results based management
RCT	Random control trial
SAVI	The State and Accountability Initiative
SDG	Sustainable Development Goals
Sida	Swedish International Development Cooperation Agency
TOC	Theory of change
TWP	Thinking and Working Politically
UNDP	United Nations Development Fund
USAID	United States Agency for International Development

Results and Evaluation Glossary⁷⁰

Attribution	The ascription of a causal link between observed (or expected to be observed) changes and a specific intervention
Baseline study	An analysis describing the situation prior to a development intervention, against which progress can be assessed or comparisons made
Blue Marble Evaluation	New initiative that plans to apply world systems thinking to evaluating global programmes
Causal inference	Conclusion that a cause is linked to an effect
Complex adaptive systems	Dynamic systems that are able to adapt in and evolve with a changing environment. There is no separation between a system and its environment: a system always adapts to a changing environment. Rather, a system is closely linked with all other related systems making up an ecosystem. Within such a context, change needs to be seen in terms of co-evolution with all other related systems, rather than as adaptation to a separate and distinct stimulus or environment
Complexity science	Complexity science is the scientific study of complex systems with many parts that interact to produce global behaviour that cannot easily be explained in terms of interactions between the individual constituent elements. Complex systems include IT networks, ecosystems, brains, markets, cities and businesses

⁷⁰ Main sources are Argyris and Schon (1974); BetterEvaluation website: <http://betterevaluation.org>; OECD. (2010). Glossary of terms in evaluation and results based management <http://www.oecd.org/development/peer-reviews/2754804.pdf>; and PowerCube <http://www.powercube.net>.

Counterfactual	The situation or condition which hypothetically may prevail for individuals, organisations or groups were there no development intervention
Developmental evaluation	An evaluation approach that can assist social innovators develop social change initiatives in complex or uncertain environments. Originators liken their approach to the role of research & development in the private sector product development process because it facilitates real-time, or close to real-time, feedback to program staff thus facilitating a continuous development loop
Evaluation design	Overarching logic for evaluations. Includes: questions, theory used to analyse data, data and use of data
Double loop learning	Double loop learning goes deeper than single loop learning and questions assumptions about strategies. It involves asking: 'Are we doing the right thing?'
Experiential learning	Learning based on direct practitioner experience or monitoring data that may not have been generated in a way that allows confident conclusions about the role an intervention has played in change
Experimental design	Evaluation design developed in the natural and medical sciences. A 'treatment' or intervention is applied to a subject or group of subjects, and observations made of what happens to subject(s) are compared (through statistical analysis) with observations of a 'control group' or counterfactual that is isolated from the intervention, e.g. as in a random control trial (RCT). In principle, because the 'subjects' of the treatment are selected randomly, the only difference between them and the control group is that the intervention has been applied to the former

Formative evaluation	Evaluation intended to improve performance, most often conducted during the implementation phase of projects or programmes
Indicator	A variable that allows the verification of changes in the development intervention or shows results relative to what was planned
Invisible power	The ways in which dominating ideologies, values and forms of behaviour are adopted without question
Logical framework	Management tool used to improve the design of interventions, most often at the project level. It involves identifying strategic elements (inputs, outputs, outcomes, impact) and their causal relationships, indicators, and the assumptions or risks that may influence success and failure. It thus facilitates planning, execution and evaluation of a development intervention
Outcome	The likely or achieved short-term and medium-term effects of an intervention's outputs
Outcome harvesting	Outcome harvesting collects ('harvests') evidence of what has changed ('outcomes') and, then, working backwards, determines whether and how an intervention has contributed to these changes
Outcome mapping (OM)	A methodology for planning, monitoring and evaluating development initiatives in order to bring about sustainable social change. At the planning stage, the process of outcome mapping helps a project team or programme be specific about the actors it intends to target, the changes it hopes to see and the strategies appropriate to achieve these. For ongoing monitoring, OM provides a set of tools to design and gather information on the results of the change process, measured in terms of the changes in behaviour, actions or

	<p>relationships that can be influenced by the team or programme.</p> <p>As an evaluation approach, OM unpacks an initiative's theory of change, provides a framework to collect data on immediate, basic changes that lead to longer, more transformative change, and allows for the plausible assessment of the initiative's contribution to results</p>
Output	The products, capital goods and services which result from a development intervention; may also include changes resulting from the intervention which are relevant to the achievement of outcomes
Most significant change (MSC)	A story-based participatory technique used to help improve programmes by including participants in data collection and analysis to enable learning to focus the direction of work towards directions explicitly valued by participants
Paradigm	A coherent and mutually supporting pattern of concepts and ontological assumptions; values and principles; methods, procedures and processes; roles and behaviours; relationships; and mindsets, orientations and predispositions
Payment by results (PBR)	A form of financing that makes payments contingent on the independent verification of results
Performance management	Advice or decisions taken on the basis of performance monitoring or measurement to 'steer' and enhance performance and results
Performance measurement	A system for assessing performance of development interventions against stated goals
Performance monitoring	A continuous process of collecting and analysing data to compare how well a project, programme or policy is being implemented against expected results

Quasi-experimental design	Similar to experimental designs but lacking random assignment to treatment or control groups. The researcher uses different criteria to select a suitable group or situation for comparison
Real-time evaluation	Evaluation that is conducted during an intervention in order to learn and adapt to enhance impact
Realist evaluation	Realist evaluation is a theory-driven evaluation set apart by explicit philosophical underpinnings. It asks ‘What works, for whom, in what respects, to what extent, in what contexts, and how?’ In order to answer that question, evaluators aim to identify the mechanisms that explain ‘how’ the outcomes were caused and the influence of context. Realist philosophy considers that an intervention works (or not) because actors make decisions in response to the intervention (or not). The ‘reasoning’ of the actors in response to the resources or opportunities provided by the intervention causes the outcomes. Context matters: firstly, it influences ‘reasoning’ and, secondly, generative mechanisms can only work if the circumstances are right. Finally, context may provide alternative explanations of the observed outcomes, and these need to be taken into account during the analysis
Result	The output, outcome or impact (intended or unintended, positive and/or negative) of a development intervention
Results agenda	Reasons for focusing on results, e.g. for accountability, learning or decision-making
Results based management	A management strategy focusing on performance and achievement of outputs, outcomes and impacts

Results framework	The programme logic that explains how the development objective is to be achieved, including causal relationships and underlying assumptions
Single loop learning	Single loop learning involves connecting a strategy for action with a result. It explores the question 'Are we doing things right?'
Summative evaluation	A study conducted at the end of an intervention (or a phase of that intervention) to determine the extent to which anticipated outcomes were produced. Summative evaluation is intended to provide information about the worth of the programme
Theory-based evaluation	An approach to evaluation that tests theories underpinning programmes
Theory of change	A theory that explains assumptions about how and why change happens. Can be used as a programme planning and evaluation tool. Used to describe causal pathways and assumptions about how change happens in projects, programmes, or society more broadly
Triangulation	Use of two or more methods or data sources to validate the same findings or results
Visible power	Visible forms of power relating to formal rules and procedures or decision-making bodies

Annex 1: Politically Smart Locally Led Case Study Summaries⁷¹

Politically smart locally led case studies

In the Congo an innovative project for disarmament, demobilisation and reintegration took an iterative approach to implementation. An external evaluation found it was significantly more successful than three other programmes implemented by the government, the UN plus NGOs, and UNDP under the World Bank's Multi-country Demobilisation and Reintegration Programme (Gillhespy and Hayman, 2011: 35–36). 10% of ex-combatants under the programme would consider rejoining a militia compared to 58% of those in other DDR programmes and 81% of those who had received no support. The programme also scored better on other indicators of sustainable reintegration, including whether ex-combatants were able to meet basic needs, had built a house, owned land or were married. Communities supported were less fearful of ex-combatants and had higher perceptions of security. 96% of ex-combatants were found to belong to cooperatives.

The EU's Forest Law Enforcement, Governance and Trade Action Plan (FLEGT) adopted a multifaceted approach to reducing illegal logging and trade in illegally logged timber by excluding illegal timber from the EU market and improving forest governance in timber-producing countries. Due to its complexity, objectives were broadly defined at the start. The team then used iterative approaches to negotiate more specific solutions through EU legislation, public procurement policies and voluntary partnership agreements with timber-producing countries. A study of illegal logging in Republic of Congo (Lawson, 2014) found that the FLEGT VPA had contributed to legislative and rights changes. In Ghana, FLEGT contributed to forest laws that provided the basis for civil society groups to successfully challenge ministers making discretionary awards of such contracts through the issue of special permits. A 2010 Chatham House report found that the programme had a role in dramatic declines in illegal logging in countries including Brazil, Cameroon and Indonesia. Moreover, FLEGT was judged to have achieved more than previous top-down, financial and technical assistance strategies. DFID's Indonesia

⁷¹ Readers interested in accessing the evaluations cited in the Box above should consult the report by Booth and Unsworth (2014).

Tropical Forest Management Programme (1991–1999) was curtailed in 1999 due to poor results. Similarly, a DFID forestry project in Ghana, prior to FLEGT, failed as it did not engage with the underlying political economy dynamics in the forest sector.

The Pyoe Pin in Myanmar/Burma (phase 1 budget £4m, phase 2 £12.8m) was initiated in a challenging, unstable political environment in which donors had to use locally led, iterative approaches to identify entry points, and to adapt to rapidly changing circumstances. It did not address a specifically defined problem but instead explored ways to support nascent civil society groups using a salient ‘issue’ approach that made it easier to develop relationships and trust with government and civil society actors. Reviews found that the facilitation of ‘issue based’ networks and coalitions and practical help enabled improved rice marketing, policy changes relating to HIV and improved quality of monastic education. Less tangible outcomes included building social capital as a step towards building political capital to support democratic transition. Results identified in a mid-term review included more confident civil society actors; however, concerns were raised that increasing donor interest accompanied by large amounts of aid could undermine success.

The Western Odisha (India) rural livelihoods programme (WORLP) was a ten-year DFID-supported project (budget £32.75m), ending in 2011. It took an iterative approach to a relatively well-defined problem. An independent evaluation by ICAI two years after support ended found significant improvements in rural livelihoods, including increased income from micro-enterprises and improved land and water management. This was striking in a particularly poor part of one of the poorest states in India. Results, which included improving water resources, agriculture and income, were significantly better than results from previous programmes, and areas where the project had not been implemented or only partly implemented. Factors influencing success included: DFID’s initial ten-year commitment and long-standing involvement in Orissa/Odisha; learning from previous projects; an iterative, participatory design process; and the nurturing and prioritisation of local ownership at all levels.

Land titling in the Philippines. With funding from USAID, the Asia Foundation identified and supported a team of local ‘development entrepreneurs’ to find technically sound, politically feasible approaches to improving land rights security. The problem was well defined, but the approach was iterative and results included legislative change that achieved impressive results – a 14-fold increase in the rate of residential land titling with potential for wider economic and social benefits

(Chikiamco and Fabella, 2011; Booth, 2014). Other potential benefits include increasing incentives for household investment in property, an improved residential property market, and increased tax revenues. Social and political benefits included reduced social friction and vulnerability of households with insecure titles. The development entrepreneur approach has been much more successful (and better value for money) than a parallel and more comprehensive reform programme costing over \$38 million dollars supported by the World Bank and AusAID that was wound up prematurely.

Tax and health reform in the Philippines (the sin tax). Another programme implemented by the Asia Foundation, with flexible funding from USAID and AusAID, identified and supported a self-starting team of local reformers to tackle a steady decline in revenue from excise taxes on alcohol and tobacco. As with the land titling programme, the Foundation used a politically informed iterative process. The reform team coordinated a broad coalition of interests and, in the face of strong opposition, secured the passage in 2012 of Republic Act 10351. This closed loopholes in previous legislation leading to an 85.6% increase in revenue from the relevant excise taxes in 2013. Nearly 80% of it was earmarked to subsidising health insurance for poor people (Sidel, 2014; Booth 2014). The intervention was far more successful than similar but less politically savvy initiatives undertaken in comparable political circumstances (i.e., a reformer in the presidency).

The Enabling State Programme in Nepal was a 13-year, DFID-supported programme with a budget of £33million. It sought to address issues of weak governance and social and political exclusion that research had identified as underlying causes of conflict and poor development outcomes. It took a long-term perspective, working through a team of politically astute, well-connected Nepali staff and providing very flexible funding to a wide range of groups, including many previously unreachable by aid programmes. It achieved tangible benefits for poor people and helped shape specific government policies as well as contributing to more inclusive electoral arrangements. It is also widely recognised as having made a significant but less tangible contribution to shaping a more inclusive political system (Brown *et al.*, 2013).

Adapted from Booth and Unsworth, 2014: pages 6–12

Annex 2: Useful Reading

Chambers, R. (2010). Paradigms, poverty and adaptive pluralism. IDS Working Paper 334. Brighton, UK: Institute of Development Studies, [Online], Available: <http://mobile.opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/493/Wp344%20web.pdf?sequence=1> [4 April 2016]

This paper is Chambers' personal reflection on problems and opportunities provided by major developments between 1995 and 2010, which he considers in terms of a new paradigm. First, the accelerating nature of change, most obviously in communication technology and Web 2.0, but also changes in the conditions, awareness, priorities and aspirations of those who were considered marginal and vulnerable and live in poverty. Second, a shift from the more participatory and permissive approaches of the 1990s to a more control oriented upward accountability, as with results based management and 'rigorous' impact assessment. Third, in contrast and conflict with that trend and largely unnoticed, a rise in the use of participatory methodologies (PMs). Fourth, theoretical understandings of the nature of technology and of complexity provide lenses, language and insights for understanding and interpreting the ontology of development. And fifth, the significance of power and relationships, including interpersonal power and relationships, has become more visible and acknowledged. It includes some simple but illuminating diagrams relating to some of the assumptions discussed in this paper.

Eyben, R. (2013). Uncovering the politics of 'evidence' and 'results'. A framing paper for development practitioners. Brighton, UK: Institute of Development Studies, [Online], Available: <http://bigpushforward.net/wp-content/uploads/2011/01/The-politics-of-evidence-11-April-20133.pdf> [April 4 2016]

This paper is a practitioner's guide to the politics of evidence and results that was written as a background paper for the Politics of Evidence conference organised by the Big Push Forward in 2013. The paper seeks to answer frequently asked questions by development practitioners concerned about the results agenda. These include:

- Why and under what conditions do potentially useful approaches – such as theories of change – mutate into coercive instruments?
- Where and how did the results based management and evidence-based policy and programming originate?
- Why are ‘results’ and ‘evidence’ discourses increasingly influential, particularly in the international development sector?
- Which organisational actors and interactions are promoting these discourses?
- Who is coming under pressure and how are some people resisting?
- What are the effects of the power of these discourses on transformative agendas?
- What are the possibilities to create spaces for different approaches?

It starts by looking at the ‘institutional artefacts’ – rules and procedures – that translate the discourse of results and evidence to reality of practice. It then unpacks ‘results’ and ‘evidence’ and finds strong family resemblance between the two discourses. These share a certain way of knowing the world (epistemology), including assumptions that evidence pertains to verifiable and measurable facts as categories of things and a particular understanding of causality, efficiency and accountability. Eyben then goes on to look at how and why these discourses have played out in the development sector, including what and who is driving the agenda.

Gulrajani, N., & Honig, D. (2016). Reforming donors in fragile states. Report. London, UK: Overseas Development Institute, [Online], Available: <https://www.odi.org/sites/odi.org.uk/files/resource-documents/10479.pdf>. [Accessed 1 June 2016]

This paper identifies ways in which donors can be more effective in fragile and conflict-affected states by exploiting theories and concepts drawn from public management. Fragile contexts demand donors look beyond blueprint solutions and work with greater sensitivity to local environments, adaptation to local contexts and enhanced organisational flexibility. Public management theory can help donors organise themselves to put these principles into practice. This paper hopes to widen the options donors have at their disposal to support

organisational reform and advance more effective ways of working in fragile states.

Hummelbrunner, R. and Jones, H. (2013b) A Guide to Managing in the Face of Complexity. London: Overseas Development Institute, [Online], Available: <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/8662.pdf> [April 4 2013]

Complexity heightens the importance of effective management, but poses challenges for the tools and approaches used most widely in international development. This guide provides an overview of these challenges and proposes a way forward:

- Management tools need to be chosen to match the situation in hand, based on whether capacities are distributed, goals are divergent, and whether there is considerable uncertainty.
- Managing in the face of complexity should be guided by three key principles: decentralised, collaborative and adaptive management.
- A selection of appropriate approaches illustrates how these principles can be applied in practice.

At the end, the guide provides readers with further resources on the subject.

Root, H., Jones, H. and Wild, L. (2015) Managing Complexity and Uncertainty in Development Policy and Practice. London, UK: Overseas Development Institute.
<http://www.odi.org/sites/odi.org.uk/files/odi-assets/events-documents/5191.pdf> [3 April 2016]

This report draws together the thinking around some of the main elements of working with uncertainty, focusing on the development of collective action, the role of networks and coalitions, and the specific features of change itself. Borrowing from elements of complexity thinking, political economy analysis and more, it elicits insights about how change happens. It also highlights some tools that can provide a more dynamic analysis of change processes. To embrace a more mature phase of working with adaptive, complex systems and changing incentives will require some very specific amendments to the dominant approaches in planning, programme design and monitoring.

Previous EBA-reports

2016:06 *Vem beslutar om svenska biståndsmedel? En översikt*,
Expertgruppen för biståndsanalys

2016:05 *Pathways to change: Evaluating development interventions with
Qualitative Comparative Analysis (QCA)*, Barbara Befani

2016:04 *Swedish responsibility and the United Nations Sustainable
Development Goals*, Magdalena Bexell, Kristina Jönsson

2016:03 *Capturing complexity and context: evaluating aid to education*
Joel Samoff, Jane Leer, Michelle Reddy

2016:02 *Education in developing countries what policies and
programmes affect learning and time in school?* Amy Damon, Paul
Glewwe, Suzanne Wisniewski, Bixuan Sun

2016:01 *Support to regional cooperation and integration in Africa – what
works and why?* Fredrik Söderbaum, Therese Brolin

2015:09 *In search of double dividends from climate change interventions
evidence from forest conservation and household energy transitions* G.
Köhlin, S.K. Pattanayak, E. Sills, E. Mattsson, M. Ostwald, A. Salas,
and D. Ternald

2015:08 *Business and human rights in development cooperation – has
Sweden incorporated the UN guiding principles?* Rasmus Klocker
Larsen, Sandra Atler

2015:07, *Making development work: the quality of government approach*,
Bo Rothstein and Marcus Tannenber

2015:06, *Now open for business: joint development initiatives between
the private and public sectors in development cooperation*, Sara
Johansson de Silva, Ari Kokko and Hanna Norberg

2015:05, *Has Sweden injected realism into public financial management
reforms in partner countries?* Matt Andrews

- 2015:04, *Youth, entrepreneurship and development*, Kjetil Bjorvatn
- 2015:03, *Concentration difficulties? An analysis of Swedish aid proliferation*, Rune Jansen Hagen
- 2015:02, *Utvärdering av svenskt bistånd – en kartläggning*, Expertgruppen för biståndsanalys
- 2015:01, *Rethinking Civil Society and Support for Democracy*, Richard Youngs
- 2014:05, *Svenskt statligt internationellt bistånd i Sverige: en översikt*, Expertgruppen för biståndsanalys
- 2014:04, *The African Development Bank: ready to face the challenges of a changing Africa?* Christopher Humphrey
- 2014:03, *International party assistance – what do we know about the effects?* Lars Svåsand
- 2014:02, *Sweden's development assistance for health – policy options to support the global health 2035 goals*, Gavin Yamey, Helen Saxenian, Robert Hecht, Jesper Sundewall and Dean Jamison
- 2014:01, *Randomized controlled trials: strengths, weaknesses and policy relevance*, Anders Olofsgård