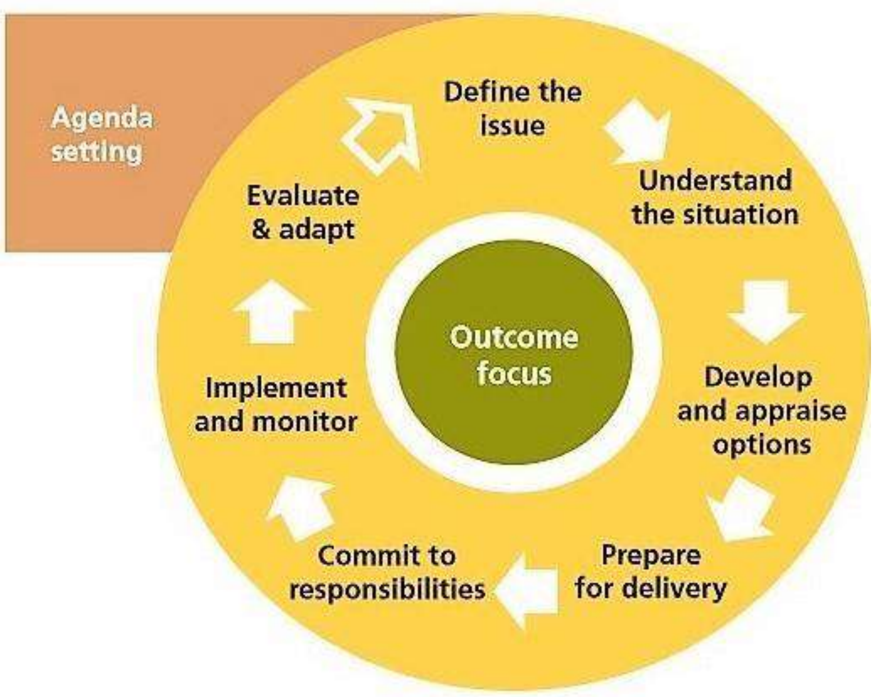
Fellowship Report

How to improve the evaluation of complex systems to better inform policymaking

Learning from evaluating Defra's Reward & Recognition Fund



Sara Giorgi

psi Policy Studies Institute
BROOKLYNDHURST
cecan Centre for the Evaluation of Complexity Across the Nexus

Executive Summary

Improving the quantity and quality of the materials recycled and reused is a key part of moving towards a 'zero waste' and circular economy. In a bid to make recycling rates go further and faster, in 2011 Defra launched the Reward and Recognition Fund (RRF). This programme aimed to explore new approaches for rewarding and recognising people for adopting positive waste behaviours focusing on food waste, recycling, reuse, waste prevention, and waste reduction. From 2011 to 2014, up to £2 million of funding was made available to 31 pilots led by local authorities, community organisations and partnerships. All schemes aimed to engage and encourage people to recycle and reuse, using rewards and recognition. The schemes, however, were very diverse, focusing on different: behaviours, geographical locations, population reach, target audiences, delivery mechanisms, engagement approaches, waste material type (e.g. food, reuse, dry recycling and specific target materials), reward details, time frames, etc.

The evaluation of the RRF focused on the outcomes and impacts of the schemes individually and the programme as a whole. The evaluation, however, did not address the experience of both practitioners and policymakers in the evaluation process. This research fills that gap and attempts to bridge the divide between practitioner and policy stakeholders by generating insights which inform the development of 'real world' evaluation methods and policy practices.

Moving beyond the lessons learned from the RRF, this research offers insights into the challenges of evaluation of complex issues and systems, and how evaluation and complexity can best inform policymaking. Research which gives a voice to local practitioners, as well as policy stakeholders, who are often funders of local projects and services, is in short supply. Furthermore, it is rare that these discussions - on how bettering evaluation of complex systems can, in turn, improve policymaking - give fruit to practical guidance.

Using the RRF has a case study, this research had three main objectives:

- To explore practitioner and policy stakeholders' evaluation experiences and challenges;
- To delve into the dynamics between evaluation and policymaking with regard to the policy cycle; and
- To investigate how complexity is understood by policy stakeholders and practitioners and how it could be useful in policymaking.

This research is mainly based on the findings of twenty semi-structured interviews with ten practitioners delivering local projects and services, and ten central Government policymakers including policy advisors, social researchers, evidence managers and strategy leads. A secondary review of site visit and workshop notes collected in the original RRF evaluation was also conducted and informed the research findings presented in this report. Research insights were presented and considered in a workshop with researchers, policymakers and local practitioners – emerging discussion points and suggested tasks (mainly for CECAN) from this workshop also informed this research.

The findings are arranged around three themes: experience of evaluation; informing policy and understanding complexity. Throughout, where relevant, the views of practitioners and policymakers are compared and contrasted.

From the interviews, there was a call for evaluation to be an integrative, continuous process not a one-off exercise at the end or a series of self-contained steps; evaluation needs to become a way of working. In the

experience of practitioners, evaluation and monitoring should be embedded in the delivery process of projects - where possible with a dedicated resource responsible for the evaluation and its outcomes.

When it comes to informing policy, interviewees suggested that the time scale disparity between policy and evaluation needed to be openly acknowledged. Both researchers and policy leads need to feel comfortable with 'good enough' and 'at the time' insights taking into account their associated risks and limitations, this would enable a better feedback loop from evaluation to policy. The policy cycle had currency and was thought to be a valuable description of how policy is or ought to be developed. Interviewees made a call for embedding 'preparing for evaluation' and 'evidence gathering' throughout the policy cycle especially in the initial policy design phase. There was a desire to ensure that the policy cycle is a shared and commonly owned tool. Policymakers, delivery practitioners and researchers should interact and work together in the formulation of policy, especially in the spirit of open policymaking and policy innovation.

With regard to understanding complexity, interviewees expressed a need for a clear definition of complexity which includes a distinction between complex and complicated as the language of complexity can be/is heard differently by different audiences. The label of complexity should not be used to deter action on important policy issues or dissuade evaluation but recognising complexity explicitly can be an opportunity and better equip policy stakeholders and practitioners with the 'smart' (i.e. fit for purpose, lean, appropriate, robust) evaluation approaches. Complexity tends to merge appraisal and evaluation in policy, conflating learning with evaluation as the evaluation of complex issues needs to be flexible and adaptive.

CECAN can further enrich the evidence base and build capacity around these three themes by improving evaluation of complex issues to better inform policy-making. Specifically, there was a request for CECAN to organise more interdisciplinary events and exchanges which encourage co-design, co-production and co-working between researchers, practitioners and policymakers. There was also a desire for CECAN to generate 'smart', complexity-appropriate methods which withstand scrutiny from all parties (e.g. economists, Treasury, etc.) and are sensitive to the current times of austerity. Finally and more specifically, there was a request for CECAN to develop a guide explaining the concept of complexity for local delivery practitioners.

Table of Contents

Introduction to research

This introduction provides a synopsis of Defra's Reward & Recognition Fund; outlines the research objectives; explores why there is a need for this research; and presents the purpose and structure of this report.

Defra's Reward & Recognition Fund and other reward initiatives

Recycling rates across England have gradually increased towards the 2020 50% target,[1] however, looking at the general trend, in more recent years, recycling rates seem to be plateauing. In a bid to push recycling rates 'further and faster', the 2010 Coalition Government wanted to explore rewarding households for doing the right thing as a mechanism to increase recycling.

As part of the Waste Review[2], in June 2011, Defra launched the Reward and Recognition Fund (hereafter RRF). This programme explored a range of approaches for rewarding and recognising people for adopting positive behaviours towards managing their waste. Up to £2 million pounds of funding and support was made available. Schemes led by local authorities, civil society/community organisations and partnerships were invited to apply. In total 31 different schemes were funded. All schemes aimed to engage and encourage people to recycle and reuse through individual prize draws, individual rewards, community rewards and/or recognition (i.e. feedback). All schemes focused on various and different: behaviours, geographical locations, reach, target audiences, delivery mechanisms, engagement approaches, waste material type (e.g. food, reuse, dry recycling and specific target materials), reward details, time frames, etc. The intention behind Defra's RRF was to create a safe space to foster diverse and innovative schemes which would generate learning about what works and what does not work, and, in turn, inform best practice. Alongside this programme, a process and narrative evaluation using key impact indicators and case studies was carried out by Brook Lyndhurst.[3]

By way of background, in 2011, the Policy Studies Institute was commissioned by Defra to undertake a 'Review of evidence on the use of reward and recognition schemes in enhancing recycling and waste prevention behaviours'. The research identified a number of evidence gaps, namely:

- The need for more rigorous and long-term approaches to evaluation in order to understand the specific mechanisms by which rewards and recognition work with diverse target groups;
- The cost effectiveness of reward and recognition schemes;
- The influence of rewards on consumption i.e. potential rebound effects; and
- The long-term effects of rewards and recognition on behaviour and measures needed to sustain behaviour change.[4]

[1] Article 11(2) of Directive 2008/98/EC – Waste Framework Directive: Targets and Reporting. http://ec.europa.eu/environment/waste/framework/targets.htm
[2] Defra (2011). Government Review of Waste Policy in England 2011. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/69401/pb13540-waste-policy-review110614.pdf
[3] Brook Lyndhurst for Defra (2016). Evaluation of the Waste Reward and Recognition Scheme http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&ProjectID=18513
[4] Bell, S., McGeevor, K., Mocca, E. and Shaw, B., Policy Studies Institute for Defra - (2013) 'Synthesis Report: Review of evidence on the use of reward and recognition schemes in enhancing recycling and waste prevention behaviours (EV0528)' http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=17819#Description

The impacts of UK recycling incentive schemes have also been assessed by research commissioned by Serco and undertaken by Eunomia.[5] Although the evidence base was not large enough to allow for robust conclusions, this work found that, whilst reward schemes in some cases appeared to be making a positive contribution to recycling performance, the results achieved by reward schemes were highly variable.

Recent research from the Waste & Resources Action Programme (WRAP) states that local authority controlled conditions (e.g. range of materials targeted and bin capacity), as opposed to socio-economic or regional variables, explain 39- 65% of the variation in recycling rates. Higher recycling rates are linked to lower effective weekly residual capacity, separate food waste and organics collections, and collection of plastic pots, tubs and trays.[6]

Research objectives

This specific research was conducted as part of a Fellowship funded by the Centre for Evaluation of Complexity across the Nexus (CECAN) with colleagues at the Policy Studies Institute and University of Surrey acting as mentors.

This research's three main objectives were:

- To explore practitioner and policy stakeholders' evaluation experiences and challenges especially those encountered in Defra's Reward & Recognition Fund;
- To delve into the dynamics between evaluation and policymaking with regard to the policy cycle; and
- To investigate how complexity is understood by policy stakeholders and practitioners, and how it could be useful in policymaking.

The above aims were unpacked producing this more detailed list of research questions:
From a practitioner[7] perspective:

- What was the practitioner experience of the Reward & Recognition Fund evaluation?
- How was the dynamic between scheme delivery and scheme evaluation?
- What were the difficulties encountered when evaluating the scheme?
- How did the experience of evaluation in the Reward & Recognition Fund compare with other programmes?
- Is the Reward & Recognition Fund complex? If so, why?

From a policy[8] perspective:

- How do the various phases of evaluation (design and planning; commissioning and detailing; implementing and analysing; and completion and using results) fit together?
- How, if at all, is evaluation used to inform policy? How was the evaluation of the Reward & Recognition Fund used in policy circles?
- How is the policy cycle used? How does evaluation fit into the policy cycle?
- Is the Reward & Recognition Fund a complex programme? Why or why not?

---

[5]Serco and Eunomia (2014) Investigating the Impact of Recycling Incentive Schemes: Full report. https://www.serco.com/Images/Serco%20Eunomia%20Incentives%20Full%20Report_tcm3-44276.pdf
[6]WRAP (2015). Analysis of recycling performance and waste arisings in the UK 2012/13. http://www.wrap.org.uk/content/factors-influencing-recycling-performance
[7] See how this is defined in the section below called 'Report purpose and structure'.
[8] See how this is defined in the section below called 'Report purpose and structure'.

- Is complexity used in policymaking? Are evaluations of complex programmes commissioned/treated/implemented differently? Why or why not?

Why is there a need for this research?

The evaluation of the RRF focused on the outcomes and impacts of the individual schemes and the programme but did not address the experience of both practitioners and policymakers in this evaluation process. This research fills that gap and attempts to bridge the divide between practitioner and policy stakeholders by generating insights which inform the development of 'real world' evaluation methods and policy practices.

Moving beyond the lessons learned from the RRF, this research offers insights into the challenges of evaluation of complex issues and systems, and how evaluation and complexity can best inform policymaking. Research which gives a voice to local practitioners, as well as policy stakeholders, who are often funders of local projects and services, is in short supply. Furthermore, it is rare that these discussions - on how bettering evaluation of complex systems can, in turn, improve policymaking - give fruit to practical and actionable insight.

Report purpose and structure

This guidance report has two main audiences in mind: policy stakeholders and practitioners. The policy stakeholders are understood to include analysts, social researchers, operational researchers, research managers, policy advisors/leads, implementation delivery officers/managers and strategy leads that work in Government departments or agencies on policy development. The practitioners are defined as managers delivering on-the-ground, local projects and services that work for community groups, businesses, charities and local authorities. Practitioners may also have a background or role in evaluation and research, but are not necessarily evaluators. Given this audience, insights are backed up by the evidence but presented in a simple, engaging and practical manner.

This report aims to provide insight and guidance on how to improve evaluation (especially of complex issues/systems) with the aim of better informing policymaking. The views of the practitioner and policymaker interviewees, where relevant, are compared and contrasted to draw out these insights further. The RRF evaluation is used as a case study from which to draw examples but the findings and insights are not restricted to the field of waste and resources.

This guidance report has the following structure:

- Data and research method – a brief account of the data used in the analysis and the research methodology;
- Experience of evaluation – an exploration of the overall importance, journey and challenges of evaluation and an analysis of the dynamics between delivery and evaluation from a practitioner perspective;
- Informing policy – an investigation of how evaluation and policy fit together at a general level and an in-depth look at the policy cycle;
- Understanding complexity - a review of the perceived definition of complexity and its nuances, and suggestions of if/how complexity can influence policy assessment and evaluation methods; and
- Conclusions – key take-outs including main research insights and suggestions for CECAN going forward.

Each of the three middle chapters (Experience of evaluation; Informing policy; and Understanding complexity) end with a text box which highlights particular aspects of the RRF case study.

Data and research method\

This section presents the data and the research methodology used in this study.

Data

Data from four sources were used:

- Notes on the programme-level monitoring and evaluation from site visits with the 31 RRF scheme managers;
- Notes on evaluation challenges and experiences from two workshops ran as part of the evaluation of the RRF;
- Ten interviews with policy stakeholders working at the crossroads of food, energy, water and environmental issues;
- Ten interviews with practitioners delivering environmental projects on-the-ground in local communities; and
- Discussions from the final formal workshop on research insights held in March 2017 with 27 delegates: 10 from policy - three government departments and one former government agency/charity represented; 7 practitioners from charities, community groups and local authorities; and 10 researchers from consultancies and universities.

The first two sources were from the original evaluation of Defra's RRF[9] but had not been analysed or incorporated fully in the published report. The last three sources are from primary research. The interviews and workshop were generated as part of this research fellowship (see methodology section). The bulk of the analysis and, therefore, research findings is drawn from the primary research, while the secondary data review was used to plug any evidence gaps and to sense check findings.

All findings and derived insights come from the perspective of the interviewees and workshop attendees and, where relevant, RRF scheme managers from the site visit notes and workshop attendees from the workshop notes. This means that all findings and insights are 'perceived' to be so from the view of the participant.

Methodology

The bulk of the findings are derived from the two sets of interviews. The interview process followed these steps:

- Development of semi-structured topic guide with questions grouped under key headings;
- Design of key selection criteria for interviewee sample;
- Interview set-up and circulation of participant information sheets to interviewees;
- Design and development of thematic, analytical data capture log in Excel;
- Interview conducted and recorded with interviewer taking notes; and
- Interview transcribed and interview notes written up in thematic, analytical data capture log.

The ten practitioner interviews were sourced from the 31 RRF schemes ran by 25 different organisations. In addition to the organisation still being in existence and the interviewee still being involved in the organisation, the key selection criteria were a range of:

- Experience of evaluation with other programmes and funds;
- Experience of various evaluation issues;

---

[9] Brook Lyndhurst for Defra (2016). Evaluation of the Waste Reward and Recognition Scheme http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&ProjectID=18513

- Scheme impacts from low to high;
- Evaluation types (in-house or contracted out);
- Local Authorities, charities and community groups; and
- Interests in evaluation learning.

The ten policy stakeholder interviews were sourced from policy contacts from the delivery of the RRF evaluation and others in research, policy development, policy implementation, and strategy across Government departments and agencies working at the intersection of food, energy, water and environment issues. It was an opportunistic sampling approach based on the researchers' own connections, snowballing and responsiveness/availability of interviewee candidates. The key criteria used were:

- Range of involvement in RRF from 'very involved' to 'heard about it but not (that) involved';
- Mixture of social researchers, evaluation, strategy, policy and delivery representatives;
- All involved in developing and/or delivering policy; and
- Having an interest or knowledge in evaluation, complexity, waste & resources or circular economy.

The semi-structured telephone interviews lasted approximately 45 minutes and were recorded and then transcribed. After the interviews were conducted, notes were written up in an analytical data capture log in Excel. This log had key themes and emerging issues drawn out from the research questions and topic guide questions under which notes were recorded. Markers for relevant quotes which backed-up an emerging theme or point were logged as well. Throughout the interview process and analysis, data were anonymised, and data protection and confidentiality assured. The site visit programme-level monitoring and evaluation approach notes were analysed in the same manner. The workshop notes were used more to sense check emerging findings.

Emerging findings were discussed and refined via brainstorms with Fellowship mentors. Insights and findings were presented at a formal final workshop which gathered delegates from three fields:

- practitioners who deliver services and projects to local communities,
- evaluators and researchers, and
- policy leads from the research, delivery and strategy perspective.

In addition to presenting research findings, a constructive dialogue and a knowledge and practice exchange amongst policy stakeholders, practitioners and researchers was facilitated and informally recorded. These notes have also informed the conclusions of this guidance report.

Emerging findings were grouped under three main themes: experience of evaluation; informing policy and understanding complexity, hence the structure of this report. Anonymised quotes pepper this report with the intent of bringing the research to life and providing evidence for the insights.

Given the very small sample size of this research, its value is in its rich, in-depth qualitative nature and is not representative of the wider population of the two interviewee groups. Given the research questions focus on exploring individual experiences and views, and the need to delve deeper into the motivations of 'why' and 'how' things are how they are, a qualitative research method was deemed most appropriate. As stated by Hakim, *"the great strength of qualitative research is the validity of the data obtained: individuals are interviewed in sufficient detail for the results to be taken as true, correct, complete and believable reports of* *their views and experiences…the other great strength of qualitative research is in the study of motivations and other connections between factors.'*'[10]

The main type of bias from the participants' perspective is acquiescence bias; where a respondent is inclined to agree with and be positive about what the interviewer presents.[11] This was mitigated through the use of carefully worded questions that do not imply there is a 'right answer' (and often given the subject matter there wasn't) and by following-up with further questions to sense check responses. From the researcher's point of view the main sources of bias were confirmation bias when a researcher has a preconceived hypothesis and uses participants' responses to validate this belief and leading questions and wording bias.[12] These biases were mitigated through ensuring a genuine open, thematic analytical process guided by the research questions and emerging themes not predetermined topics, and that interview questions were not leading. It is standard practice for the researcher to repeat back the answer to double check understanding and give respondents an opportunity to ensure that what they said was correctly understood and interpreted. It needs to be acknowledged that, though care and rigour has been applied to the analysis, given the nature of the research and methodology there will be an element of judgement and interpretation inherent in the analysis.[13]

Experience of evaluation

This section explores the importance, journey and challenges of evaluation from the perspective of the research participants. It also investigates and unpacks the relationship between delivery and evaluation. This chapter concludes by summarising why these insights into evaluation matter and how they can improve policymaking.

Importance of evaluation

Evaluation was deemed important for both practitioners and policy stakeholder interviewees. Evaluation was considered instrumental in measuring the effectiveness and impact of, on the one hand, local projects and, on the other, policies. Amongst respondents there was a sense of realism that, at times, circumstances and data do not allow for evaluation outcomes to influence policies.

Practitioner interviewees felt the strain of providing accurate data quickly which was juxtaposed with the 'messy' nature of delivery and the slowness of behaviour change. The need, however, for demonstrating scheme effectiveness was appreciated. Practitioners also mentioned that it was useful and important to feel part of a

"Certainly being part of [and] knowing that it's contributed to a wider sort of base of knowledge I felt was really beneficial. Our council has always been keen to be part of the sort of national picture, we sort of strive to be, to improve as much as we can…"
Practitioner

"I think evaluation is very important and I think it is sometimes overlooked. It is very easy to go and say "Go and deliver this, here is a bit of money", but you never quite understand whether something is successful or not, whether that action is successful or not, and it achieves the ripples that you need it to achieve, without proper evaluation. Now there are degrees of evaluation…"
Policy stakeholder

[10] Hakim, Catherine (2000). Research Design: Successful designs for social and economic research. Second Edition. Routledge: London and New York, p. 36.
[11] DeVellis, Robert F. (2017). Scale Development: Theory and Applications. Fourth Edition. Sage: Los Angeles.
[12] Nickerson, Raymond S. (1998). 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises.' Review of General Psychology, Vol. 2 No. 2, pp 175-220.
[13] Sarmiak, Rebba (August 2005). '9 types of research bias and how to avoid them.' Quirk's Marketing Research Review. www.quirks.com/articles/9-types-of-research-bias-and-how-to-avoid-them

national programme and contribute to the wider evidence base on reward and recognition.

Policy stakeholders all stated that evaluation was and ought to be key to good and open policymaking. Evaluation, often indirectly and in certain circumstances, was believed to inform policy; however, interviewees stated that policy is not driven by evaluation outcomes. Often policy interviewees highlighted two facades of evaluation: an external, formal, independent assessment carried out for accountability purposes and an internal more iterative and reflective dialogue of what works. This conflation of assessment and learning in evaluation alimented the call for evaluation to be clearly defined from the onset, fit-for-purpose, lean and adaptive or, effectively, '*smart'*. The data collected suggests that there is a need for 'smart', complexity-appropriate methods which withstand scrutiny from all parties (e.g. economists, Treasury, randomised control trial supporters, etc.) and are sensitive to current policy timescales and times of austerity.

Evaluation journey – a continuous process or a collection of independent steps?

The diagram above depicts the different phases of an evaluation.[14] Policy interviewees were asked about the different phases and if and how they fit together. The overriding consideration was that the phases linked up in theory, while, in practice they are a collection of self-contained steps. Interviewees recognised that as evaluation is meant to be done as a continuous process, having phases operate in isolation or in silos is not conducive to better evaluation practice. One of the issues was that often each phase was 'owned' and managed by different people or teams. A specific criticism of the 'design & plan' phase was that it was often left up to the research community to address this in the tendering process rather than the commissioning Government department deliberating and specifying the evaluation plan beforehand.

"What we tend to do is produce a huge wish list of things we would like the evaluation to achieve…Everything else flows reasonably well except for the fact that we sometimes change our minds, so commissioning implementation etc. will broadly happen then in the way you would expect them to do within consultation. But I think the thing that's really missing is that actually designing and thinking logically about what we really want before we actually commission it."
Policy stakeholder

From the accounts of policy interviewees and the experience of practitioners (both from the interviews and site-visit notes), there seems to be a sliding scale of approaches from more locally-run, laissez-faire approaches of some departments to more hands-on, centrally controlled styles of evaluation of others with others falling in between these two poles.

[14] This evaluation journey was developed by the author based on her experience of conducting evaluations of policy programmes and the interviewees' descriptions but it also draws from the evaluation stages outlined in these sources: HM Treasury (April 2011). The Magenta Book: Guidance for Evaluation. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf; and United Kingdom Evaluation Society (2012) Evaluation Capabilities Framework. https://www.evaluation.org.uk/images/ukesdocs/UKES_Evaluation_Capabilities_Framework_January_2013.pdf

Evaluation challenges

This section outlines evaluation challenges, many of which were specific to the RRF but reflect broader issues true of other programme, project and policy evaluations. When it came to the challenges faced by practitioners in conducting evaluations; the interview, workshop and site visit data suggest that the three main challenges were related to:

- Data;
- Assessing impact, additionality and attribution; and
- Sourcing and setting up comparison groups.

When it came to data challenges these can be grouped into three main issues those around:

- Accessibility, availability and ownership of data;
- Data collection (e.g. knowing what data to collect, having too much data, delivery barriers impeding data collection, different reporting mechanisms etc.); and
- Data quality (e.g. the spectrum of data ranging from randomised control trial data to anecdotal evidence with everything in-between).

These and the other evaluation challenges discussed like costs, resources, capabilities and diverting time for delivery to do evaluation (see next section) reconfirm the assessment outlined in the original evaluation of the RRF.[15]

Policy stakeholders, though from a different perspective, experience the same evaluation challenges as practitioners. The main challenges mentioned by policy stakeholders were:

- Data (quality, availability and relevance);
- Isolating policy impact especially when there are many intervening factors at play in an inter-related system; and
- Identifying control groups.

Other challenges stated by the policy interviewees include: limited techniques and methods; lack of clear objectives; costs; variety and diversity of projects; and capabilities and capacity.

The relationship between delivery and evaluation

Practitioners described the dynamics between delivery and evaluation as ranging from a tense 'tug of war' to a complementary partnership.

Some practitioner interviewees had experienced a tense relationship between delivery and evaluation, with evaluation taking time, resources and energy away from delivery. Given that many practitioner interviewees worked for organisations which were focused on local service delivery, this 'tug of war' often proved problematic. Certain practitioner interviewees stated that if the information and expectations for monitoring and evaluation requirements had been made clearer upfront then this contrast could have been avoided.

"The operational side was within the service itself. So we didn't really have that conflict. ...they had ... a small number, they had a dedicated manager and some of the staff to the pilot .... So generally it was to a point that everyone was on-board and knew their individual role. ... It was our responsibility to manage the project and relay anything that needed to be done operationally to the operational team but we weren't responsible for delivering it as such."
Practitioner

[15] Brook Lyndhurst for Defra (2016). Evaluation of the Waste Reward and Recognition Scheme http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&ProjectID=18513

More often, practitioners described the relationship between delivery and evaluation as a complementary partnership. In these cases, evaluation fed into and was integrated in the scheme's delivery activities. This was often true of schemes that had allocated evaluation to a separate person, team or external organisation. Such a set-up ensured a clearly defined role and set of responsibilities providing a sense of ownership and a 'stake in' the outcomes to those running the evaluation. This mutually beneficial partnership approach seemed to work best with clear delegation, factoring in evaluation into delivery activities from the onset and, where relevant and appropriate, using specialist organisations.

Experience of evaluation | RRF Example | A special case?
The following traits relate to the interviewees' experience of the RRF evaluation, however, they offer important lessons that should be taken into account when encouraging practitioner buy-in for evaluations more generally:

- Being part of a national call to action to pilot rewards and recognition, and gather evidence on what works helps engender a sense of purpose;
- Embarking on a 'proof of concept' of whether policy has worked or not without a preconceived idea of what success looks like fosters comradery between policy and practice, (however, may hinder evaluation);
- Monitoring and evaluation expectations, be they onerous or light-touch, need to be factored in from the onset; and
- Myriads of intervening factors and background noise make attribution to the scheme, let alone to the reward element of the scheme, difficult, if not impossible.

Drawing from the all the data, the key insights around evaluation experience that can better inform policymaking are:

- A wider sense of purpose and belonging in an evaluation can help nurture buy-in;
- When evaluation is planned, resourced and budgeted from the onset, the relationship between project delivery and evaluation can be a fruitful partnership;
- When evaluation has a separate role which is 'owned' and has a stake in the outcomes, this enhances the partnership between project delivery and evaluation;
- Practitioners and policy stakeholders experience similar evaluation challenges;
- Evaluation needs to be an integrative, continuous process both in theory and practice, it cannot be perceived or delivered as a one-off standalone activity or a collection of self-contained steps; and
- At the crossroads of water, energy, food and environment, both service delivery and policy development happen in a 'messy' context with many intervening factors at play, this should be explicitly recognised.

Informing policy

This section, first, explores how evaluation and policy fit together at a general level and then takes an in-depth look at the policy cycle. This chapter concludes by highlighting why these insights can help to better integrate evaluation into policymaking.

Evaluation and policy – not a perfect match?

"The fact it [evaluation] will slow things down, it is not always possible to do that, we have to move quite fast sometimes. And evaluation has its time and place, doesn't it, and what can often happen is that you spend a lot of time doing the evaluation but you have kind of missed the opportunity…to actually make a difference was probably with six months of it being announced and you can never do an evaluation that quickly.."
Policy stakeholder

Evaluation was perceived by interviewees to be central to good policymaking and, especially, to good, evidence-backed, open policymaking. However, the extent to which evaluation managed to directly influence policy in practice was doubted by most interviewees.[16]

Policy stakeholder interviewees mentioned that one of the main issues with ensuring good, evidence-backed policymaking is informed by evaluation is timing. Policy and evaluation work at different speeds. Policy and evaluation have two distinct tempos: policy tends to be fast-paced, dynamic and with a quick turnaround while evaluation is known to have longer timeframes due it being detailed and analytical. This mismatch often means that the feedback loops between evaluation and policy do not work as planned.

With an open acknowledgement from both sides, this timing mismatch, according to interviewees, can be addressed. External evaluators, researchers and/or civil service analysts, on the one side, need to be confident in sharing emerging findings and insights live at that point in time through outlets such as learning sessions. The idea being that emerging findings are 'good enough' and tell the narrative of the 'impact at this point in time' to feed into policy development decisions. Policymakers, on the other side, need to feel comfortable with these caveats and with the risk that final results and end conclusions may differ when all the data has been collected and analysed.

A couple of policy stakeholder interviewees also mentioned that, given recent changes to the way policymaking and regulation is formulated in the UK government, evaluation is becoming more salient through mandatory post-implementation reviews and future policy reform. This means more Government officials are being exposed to evaluation tasks; therefore, disseminating evaluation practice across more colleagues.

Policy cycle

Background

A policy cycle is a tool used for planning and analysing the different phases of policy development.[17] The number and names of each phase can vary but the essence behind each step remains consistent.

Figure 1 | Defra's Policy Cycle

[16] For a definition of open policymaking, see Cabinet Office (February 2016). Open Policymaking Toolkit. https://www.gov.uk/guidance/open-policy-making-toolkit
[17] HM Treasury (April 2011). The Magenta Book: Guidance for Evaluation. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf

The general approach starts out with agenda setting which identifies the problem or issue that needs addressing. This first step often has specific phases of 'defining the issue' and 'understanding the situation'. This is then followed by steps which formulate and assess the different alternative courses of action and preparation for delivery. In the following phase, Government decides on the course of action (which includes maintaining the status quo i.e. taking no action). The decision made in the previous step will then be put into practice through implementation and monitoring. The final phase (which is the first step in the next cycle) is about assessing the effectiveness of the policy in terms of its intended objectives, outcomes and impacts. This 'assessment of effectiveness' is done through evaluation and adapting lessons learned into the future delivery of the policy.

As mentioned previously there are many versions and adaptations of the policy cycle, the one used to stimulate discussion in the interviews (shown above) has been used within Defra.[18]

How is it used?

Most policy interviewees tended to be familiar with the policy cycle. It was said that it tended to be applied to new policies and the cycle was used as a discursive process tool. Some policy interviewees had used it in presentations; often to describe or justify the Department's work.

Policy interviewees didn't think the policy cycle was something they necessarily 'used', but it was an account of how policy is developed and how things work. The policy cycle was seen as part of standard operating procedures with processes aligned to it. Several interviewees did make the point that the policy cycle described how things should work in theory rather than how they actually worked in practice. Though the steps in the cycle were depicted as neat and compact, the real world was much messier and complex.

One critique expressed by a couple of policy interviewees was that though the policy cycle was in a shape of the wheel it often felt like the process it was describing was quite linear. These interviewees often stated that, though evaluation was embedded throughout the policy cycle, having it as the final step suggested it is something you only think about at the

"I think it's just how things are done I don't think people even think consciously oh no I'm going to start using the policy if you see what I mean? That is just business as usual.

Q: *That's really ting OK. So it's not like an active process it's just … [a] subconscious process?*

Yes I think it's been around so long it has truly just become how people do things. At the beginning there was probably a transition… raising it up into people's consciousness. I'm struggling to think of anything that I've seen that would skip out steps or something like that, I just can't see it."

Policy stakeholder

"An ideal model that never actually happens in practice."

Policy stakeholder

"So there is an evaluation cycle spinning around next to each of those stages of the policy cycle, how much that evaluation cycle spins really depends on the maturity of the policy."

Policy stakeholder

"There are so many 'it depends' questions that although it is a nice diagram and it is useful for explaining there are different stages, and different things and different types of organisations might work at different stages, it is not helpful from the point of view that it doesn't even begin to demonstrate the complexity and the messiness of the real world."

Policy stakeholder

[18] Defra (undated) Inside Defra: How Defra works? p. 11 http://www.larpnet.com/downloads/insidedefra.pdf *Though undated the foreword is written by Dame Helen Ghosh and dates to her time as Permanent Under-Secretary of State for Environment, Food and Rural Affairs between November 2005 and December 2010.*

end. Other interviewees, however, felt that the policy cycle was indeed a cycle as it enabled the user to dip in at the relevant phase (depending on what is driving the policy or the policy's maturity) and then follow the steps around.

Where and how does evaluation fit in?

Given interviewees' familiarity with the policy cycle most interviewees appreciated that evaluation was not simply the last step but was happening 'behind the scenes' informally at each step. Often it was stated that 'preparation for evaluation' and 'evidence gathering' happened or should happen at each phase while the completion of a formal evaluation happened at the final step 'evaluate & adapt'. Most interviewees stated that at the very least 'preparing for evaluation' should have more of a prominent role in the first few phases of the policy cycle. At the other end of the spectrum, some interviewees stated that evaluation needs to be designed into the intervention or policy from the onset so that the policy is designed to be evaluated with a constant 'test, learn and adapt' feedback loop.

There was an unclear conflation between the desired ideal of how evaluation fits into the policy cycle and the reality of the status quo of how things work. The key points on how does/should evaluation fit into the policy cycle were:

- Evidence gathering and preparation for evaluation is going on in different ways across all phases; this is/should be adapted and fed back to the relevant phase;
- Evaluation cannot delay policy development; there needs to be a timely input with quick feedback loops;
- Co-produced working between policymakers and analysts at each phase automatically embeds evaluation in each step, challenging the view that evaluation is done separately by the evaluation team and is, therefore, something that happens to policy rather than a way of working; and
- The policy cycle is/should be a 'cycle'; the 'final' step in one wheel is the 'first' step in the next wheel ad infinitum.

Policy interviewees, in the main, appeared attached to the policy cycle and felt that, though it had its limitations, it could be adapted to be more fit for purpose in the complex, inter-related system it operates in.

Challenges to embedding evaluation into the policy cycle

Most policy interviewees highlighted that there were *challenges* to embedding evaluation into the policy cycle but no systematic *barriers* – this was an important positive framing distinction. Challenges were thought of as surmountable difficulties and there were no insuperable system level barriers. Many, if not all, of these challenges have already come up in earlier sections of this report.

> *"There is something about who wants to be the Minister, or who wants to be the senior civil servant who holds up their hand and says 'Our evaluation says this didn't work', so is there something, I don't know, I haven't seen it first hand, but I think there is something in is there really a will to know what actually worked, or do we just want to be told the good news story and let's forget all the rest of it"*
> Policy stakeholder

The meta-level challenges were policies not being designed from the onset as 'evaluable' and policies taking place in an interrelated system with a myriad of intervening factors impacting a non-linear process.

The sub-level challenges can be grouped into three categories: time; working culture; and buy-in. The challenge of time when it comes to better integrating evaluation into the policy cycle relates to not having enough time (for example, not having enough time to involve relevant colleagues or to embark on co-produced working arrangements); policy and practical needs changing over time; and the timing mismatch between policy and evaluation. The working culture set of challenges relates to:

- Silo working styles which means one person or team is responsible for one policy cycle step and has limited or no knowledge of who did the previous step or who will be doing the next step;
- Not having access to colleagues (e.g. policymakers not having access to policy analysts); and
- The high turnover of staff due to how career paths get forged.

The buy-in group of challenges relate to needing senior management buy-in - which is sometimes beyond the organisation or department in question (e.g. seeking support from Treasury, National Audit Office, Cabinet Office, etc.) and political will.

The more common challenges included: mismatch and/or lack of capacity and capabilities; data constraints; and costs, especially at a time of austerity when there is little or no scope for non-legislative evaluations.

From all the data analysed, the key insights around how evaluation can be better integrated into policymaking are:

- Explicitly acknowledge and address the time scale disparity between evaluation and policy - work with ‘best available’ and ‘at the time’ insight;
- Encourage closer collaboration and co-produced working between policymakers and analysts;
- Ensure policy cycle is an actual a way of working in practice, not only in theory;
- Allow evaluation to form part of initial thinking in policy design;
- Ensure ‘preparing for evaluation’ and ‘gathering evidence’ feeds into and across each phase but without being an additional burden; and
- Make sure evaluation does not delay or derail policy development; evaluation needs to complement and enhance policy.

Informing policy | RRF Example | A special case?

The final RRF evaluation report was not ready or published in a timely fashion, which meant that rewarding positive behaviour was no longer a topical issue when the findings were disseminated. The final results as well as the progress and emerging learning up-dates, however, had impact on:

- Social research:
  - Action based research projects were funded - where monitoring and evaluation is done by a separate agency but feeds back into delivery;
  - Provided insights on how to best set-up reward and recognition schemes in a waste context; and
  - Produced rich learning which was documented and shared amongst Defra analysts.
- Policy:
  - Informs current thinking on levers of behaviour change;
  - Used in different policy circles across waste streams (food, reuse and dry recycling); and
  - Informed thinking on rewards and how they are not considered a blanket measure to increase recycling.
- On-the-ground practice:
  - Report made available to other schemes;
  - May have dissuaded some local areas to take up rewards to increase recycling; and
  - Some practitioners felt RRF left a legacy in their local communities.

Understanding complexity

This section defines complexity from the perspective of the research participants, explores the nuances behind the label 'complexity', and offers insights on how complexity can influence policy assessment and evaluation methods.[19] This chapter concludes by emphasising how complexity can better inform evaluation and policymaking.

Defining complexity

When defining complexity the discussions with interview participants were anchored to the RRF programme, evaluation and the policy issue it was tackling. At a very broad level, there were many different nuances of complexity, however, many descriptions had come traits. Largely due to the line of questioning, definitions of 'complex' were juxtaposed to 'complicated'.[20]

> *"Comparing different schemes it's very difficult to transfer a scheme from one area to another in those terms. Sometimes communities can vary wildly from area to area so you might find your trial area works very well but if you transfer that even to a community that's next to it, it might be very difficult to replicate the results."*
> Practitioner

From the practitioner point of view, interviewees tended to have a 'local' perspective of complexity often focusing on the micro-level of scheme and service delivery. Indicators of complexity were considered to be when: there is a lot of 'background noise'; the issue addressed is new, innovative or unexplored; the issue focused on is considered controversial or contentious; the scheme brings about unintended benefits beyond the objectives set out; and the scheme cannot be easily transferred to or replicated in a new context. In addition, as practitioners were aiming to change human behaviour, quite a few mentioned that understanding how behaviour change works is always complex.

From the policy stakeholder point of view, interviewees tended to have more of a 'system' perspective of complexity often concentrating on the holistic and meta-level picture. Indicators of complexity were considered when: an intervention had a different impact across the same audience; the impact cannot be visualised (e.g. waste journey); and, most importantly, a policy is interrelated to other policies/factors making it extremely difficult to assign causality. A distinction often made by policy stakeholders was that the challenge can be complicated or simple while the system with its intervening factors, multiple interactions and trade-offs is complex.

> *"But also you're dealing with quite complex systems where there's lots of interactions, there's lots of trade-offs and things like that that can be quite challenging."*
> Policy stakeholder

Interviewees that gave a distinction between 'complex' and 'complicated' often mentioned that what makes something complex is when you cannot predicted or control the outcomes. This was mentioned by both practitioners and policy stakeholders but it was more common, and more consistent, amongst policy

---

[19] For a definition of complexity see HM Treasury (April 2011). The Magenta Book: Guidance for Evaluation. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf and Fagan-Watson, B., Watson, T., Elliott, B., Vanner, R., Shaw, B. and Morris, S. (2015). 'What range of methods is available to evaluate the implementation and impact of complex environmental policies on complex systems? A report to the Department for Environment, Food and Rural Affairs. Policy Studies Institute. Defra, London.

[20] For a distinction between complicated and complex aspects of interventions see Table 2 in Rogers (2008). *Using programme theory to evaluate complicated and complex aspects of interventions.* Evaluation, 14(1), 29-48.

stakeholders. Most interviewees, across the two camps, also felt that even if an issue or context was complex the approach to assess the scheme's effectiveness did not have to be.

When interviewees were presented with a list of attributes[21] and asked whether they applied to the RRF, the majority of interviewees across both practitioners and policy stakeholders stated that each attribute applied. Most interviewees, especially amongst practitioners, had not necessarily thought about the RRF as complex until after having heard the attributes and having considered most of them as applicable to the RRF.

'Being context specific' is an attribute which struck a chord with all practitioners, again echoing their 'local' perspective of complexity. The 'time lag of impacts' was an attribute that rang less true for both practitioners and policy stakeholders as relevant to the RRF. Fewer practitioner interviewees, compared to policy stakeholders, felt that difficulties in placing boundaries was relevant to the RRF – this may be because practitioners were considering their own local schemes while policy stakeholders had a more system-wide view of the whole programme.[22]

'Contextual specificity', 'difficulties in ascribing impact and attribution', and 'background noise' (i.e. the amount of 'noise' in the system - other intervening factors at play) were the three attributes which seemed to resonate well with interviewees and sparked discussions around the finer detail or nuances of complexity.

Nuances of complexity – what's in a name?

In the discussions around complexity and how it relates to the RRF, interviewees made a few interesting points around the framing of complexity. The understanding of complexity may differ depending on the perspective one takes. Questions like: 'what is being assessed for its complexity?' 'Is the subject matter the evaluation, policy, issue/challenge, scheme concept, etc.?' needed to be addressed. This is an important point because it relates to framing – an issue or challenge can be complex but a policy or evaluation does not have to be complex.

---

[21] The attributes were: time (impacts not apparent until much later on); dynamic system that adapts to your intervention; non-linear impacts (where impact may happen all at the beginning, or suddenly at the end of an intervention); difficulty in disentangling the impact of your intervention in a complex, messy system etc.; the amount of 'noise' in the system; difficulties in placing boundaries (e.g. around control areas); context specific (e.g. dependent on a specific audience, area, socio-demographic make-up, etc.; and emergence – bigger than the sum of the parts and multi-level impact and interactions - upwards and downward causation. Attributes drawn from: Fagan-Watson, B., Watson, T., Elliott, B., Vanner, R., Shaw, B. and Morris, S. (2015). 'What range of methods is available to evaluate the implementation and impact of complex environmental policies on complex systems? A report to the Department for Environment, Food and Rural Affairs. Policy Studies Institute. Defra, London, pp. 12-19.

[22] Out of the ten policy stakeholders, on interviewee did not answer this question. Out of these 9, two were not familiar enough with the RRF so answered this question based on whether they thought they were characteristics true of complexity generally rather than specific to the RRF.

List of attributes presented	Practitioner (n=10)	Policy stakeholder (n= 9)
Time lag of impacts	6	7
Dynamic system	7	9
Non-linear impacts	9	9
Difficulty in disentangling impact	8	9
Background noise	8	9
Difficulties in placing boundaries	6	9
Context specific	10	8
Emergence – multi-level causation	8	9

The question of the perspective and framing of complexity is linked to language use. There may be conceptual issues around the language of complexity in policy, researcher and practitioner worlds (i.e. the difference between complicated and complexity) which are worth exploring and considering. It would be helpful and advisable to have a common definition and understanding of complexity across the research, policy and practitioner communities. At the workshop organised as part of this research with practitioners, policymakers and academics, for example, CECAN was asked to develop a guide on complexity for local actors. This would be a simple, explanatory guide on complexity concepts and their relevance for community-based evaluators (e.g. NGOs, Local Authorities).

It was clear from the interviews that complex issues are not necessarily recognised as such by policymakers or practitioners. Complex policy issues are not formally labelled as 'complex'. In the policy sphere, according to interviewees, complexity is not starkly defined as per the list of attributes but there is an awareness of complexity, almost as subtext.

The label 'complex' was not thought of as carrying negative connotations. There was an appreciation and acceptance that some policies will be complex but this should not discourage their evaluation. A couple of interviewees, however, did state that though the term 'complex' isn't off-putting it may act as a deterrent to embark on policies tackling complex issues as it would automatically mean committing more resources

Complexity influencing policy and evaluation methods

If a policy or policy issue is to be assessed on its complexity, to have traction amongst civil servants, it needs to be seamlessly integrated into existing processes and resources for formulating policy. Policymakers were keen to emphasise that a process for assessing complexity must be user-friendly and it cannot be seen as an additional compliance process burden. Such an assessment needs to be integrated in existing policymaking systems and mechanisms, for example, it could be incorporated in the current impact appraisal process. Some interviewees remained unconvinced of the benefits of making 'complexity' explicit in policy development. There was, however, agreement that any assessment of policy on complexity needed to be pitched and framed as a positive opportunity.

In the experience of the policy stakeholders interviewed, complexity comes to the fore when an issue does not have a clear governance strategy in terms of which department/team/person has responsibility for it, or when a policy was deemed intrinsically interrelated to other policies/factors that it is extremely difficult to assign causality. Interviewees stated that most issues in departments like Defra are inherently complex given that it continuously deals with nonlinear processes in an interrelated system.

Policy stakeholder interviewees were asked about whether the type of evaluation which gets carried out changes if a policy is deemed complex or not. Half of interviewees stated that

"Yes it might be that for the particular complexity that it's just not possible to use one of those research designs…If you've got an area where you've just got a lot of different policies working then actually measuring the precise impact that each one has had rather than understanding the cumulative impact can be quite tricky."
Policy stakeholder

"I don't agree with the "What works" centres, I think they are fundamentally flawed because even if you can do a really robust RCT type evaluation all that will tell you is it worked in that context at that particular time, delivered in that particular way, and we know from our experience that you don't have to deviate very much from the delivery model to get a completely different result.."
Policy stakeholder

"I think that's why CECAN was set up, isn't it, in the sense that we recognise that actually our ability to evaluate these sorts of things is not particularly great, and I think that we are hoping for insights into how to do it better. So no I don't think it is, I think it is recognised as an issue, but I don't think that currently we are particularly good at doing it."
Policy stakeholder

complexity does, or at least should, affect the type of evaluation approaches used. There was some appreciation amongst interviewees that a complex system precludes the use of certain evaluation techniques.

There was a genuine interest in CECAN's work in terms of appropriate evaluation methods for complex issues/systems. Some interviewees stated that complexity cannot be seen as the only issue in deciding which policies to test or implement, there is a bigger picture of competing priorities in appraisals (e.g. costs, political appetite, risk, impact, accountability, etc.). One interviewee also questioned the usefulness of clustering policies under different categories or evaluation techniques. There is a need to ensure we have 'smart' evaluation techniques for assessing complex issues or systems but these need to be user friendly, especially given increasing pressures of time, budget and competing priorities.

Informing policy | RRF Example | A special case?

The concept of rewarding and recognising positive behaviour is simple, it is the endless permutations of the schemes which makes the programme-level context complex The evaluation carried out was not complex but a bespoke package with tried and tested methods looking at both the micro and meta level impact. The characteristics that made the programme-level system complex were:

- A variety of challenges, behaviours, audiences, stakeholders, delivery approaches, reward mechanisms, etc.;
- The rich tapestry of context specificity;
- Substantial background noise present which could only be logged rather than mitigated;
- Too many intervening factors to establish causality and attribution; and
- An open, unbounded and interrelated system with non-linear impacts.

From all the data analysed, the key insights on how can understanding complexity better inform evaluation and policymaking are:

- Complexity can be that common trait across policy issues that have governance problems, that are interrelated, and for which impacts are difficult to measure and attribute;
- Context, attribution and background noise were aspects that resonated well with interviewees when discussing complexity;
- The label 'complexity' is not that important but its framing is. Assessing complexity in policy needs to be presented as a beneficial opportunity;
- Acknowledging complexity explicitly can improve evaluation and thus better inform policy-making;
- Any assessment of a policy's complexity has to be integrated in existing appraisal mechanisms, there is no appetite or capacity for an additional compliance process amongst civil servants; and
- There is a need for knowing which evaluation methods are best to use with complex issues and systems, and why.

Conclusions

This section highlights the main conclusions across the three themes: experience of evaluation; informing policy and understanding complexity. It then ends with suggestions for key tasks for CECAN informed by both this research study and delegates who attended the research insights workshop.

Experience of evaluation

The main take-outs around experience of evaluation from interviewees are:

- Evaluation needs to be an integrative, continuous process not a one-off exercise at the end or a series of self-contained steps, it needs to become a way of working;
- Evaluation outcomes backed-up with evidence needs to inform policy design;
- Evaluation and monitoring should be embedded in the delivery process of projects - where possible with a dedicated resource responsible for the evaluation and its outcomes and where this is not possible with sensitivity to the capacity of the people doing the delivery;
- Evaluation is not just about measuring a change in behaviour due to a new policy but is also about maintaining policies and building on what is already known to work; and
- There is a much wider debate about what evaluation is and means to different people within the policy sphere. For example, it does not mean the same thing to a social researcher, operational researcher, statistician, economist, policy lead or delivery practitioner. These people use various levels and types of evaluation; ranging from a reflective, learning process to performance indicators for monitoring. The important element is to have realistic expectations about what can be learned and achieved given the policy or programme's design.

Informing policy

The main take-outs around informing policy are:

- Acknowledge the time scale disparity between policy and evaluation, use 'good enough' and 'at the time' insights taking into account the related caveats, risks and limitations;
- Embed 'preparing for evaluation' and 'evidence gathering' throughout the policy cycle especially in the initial policy design phase;
- The policy cycle is not the sole responsibility of the policymakers, it needs to be shared and owned across all relevant parties. Policymakers, delivery practitioners and researchers should interact and work together especially in the spirit of open policymaking and policy innovation;
- Need to revisit, self-evaluate and understand to what extent evidence-backed policy and open policymaking actually happen in reality; embark on open policymaking[23] and evidence-backed policymaking; and
- There is a wider policy debate which calls for a rethinking of policy formulation to ensure that people and systems are central to policymaking, especially with appropriate levels of management so that a system with multi-level interactions can attempt to address the disparity in responsibility between centralised policymaking and on-the-ground delivery.

Understanding complexity

The main take-outs around understanding complexity are:

[23] Cabinet Office (February 2016). Open Policymaking Toolkit. https://www.gov.uk/guidance/open-policy-making-toolkit/getting-started-with-open-policy-making#co-design-policy

- There is a need for a clear definition of complexity which includes a distinction between complex and complicated as this is currently not understood in the same way by all (e.g. commissioners of evaluation and delivery organisations). The language of complexity can be/is heard differently by different audiences;
- It is unrealistic to expect to control or predict complex issues as this goes against their very nature but they can be guided and steered;
- The label of complexity should not be used to deter action on important policy issues or dissuade evaluation but recognising complexity explicitly can be an opportunity and can better equip policy stakeholders and practitioners with the ‘smart’ (i.e. fit for purpose, lean, appropriate, robust) evaluation approaches;
- Expectations of what can be achieved through evaluation of complex issues need to be managed and remain within the original scope of the evaluation, however, evaluation of complex issues with ‘smart’ evaluation approaches is worthwhile;
- Complexity merges appraisal and evaluation in policy, conflating learning with evaluation as the evaluation of complex issues needs to be flexible and adaptive; and
- How and why complexity is useful for policymaking needs to be made clearer by addressing how different approaches to thinking about complexity change the policymaking process.

Suggestions for tasks for CECAN

Building on the key take-outs from this research and those voiced by delegates at the research insights workshop, the suggestions for the next tasks for CECAN are:

- Keep doing more of the same, especially with regard to organising more interdisciplinary events to better integrate co-design and co-production working practices amongst researchers, practitioners and policymakers;
- Influence policymakers with regards to the types of evaluations conducted and help create more realistic expectations about what evaluation can and cannot achieve;
- Generate ‘smart’, complexity-appropriate methods which withstand scrutiny from all parties (e.g. economists, Treasury, etc.) and are sensitive to current times of austerity;
- Produce a simple explanatory guide on complexity concepts for community-based evaluators (e.g. NGOs, Local Authorities); and
- Update the definition of evaluation so that it is current, has mileage across and is reflective of the needs of researchers, policymakers and delivery practitioners which operate at the crossroads of environment, water, food and energy issues.

References

Article 11(2) of Directive 2008/98/EC – Waste Framework Directive: Targets and Reporting. http://ec.europa.eu/environment/waste/framework/targets.htm

Bell, S., McGeevor, K., Mocca, E. and Shaw, B., Policy Studies Institute for Defra - (2013) 'Synthesis Report: Review of evidence on the use of reward and recognition schemes in enhancing recycling and waste prevention behaviours (EV0528)' http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=17819#Description

Brook Lyndhurst for Defra (2016). Evaluation of the Waste Reward and Recognition Scheme http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&ProjectID=18513

Cabinet Office (February 2016). Open Policymaking Toolkit. https://www.gov.uk/guidance/open-policy-making-toolkit/getting-started-with-open-policy-making#co-design-policy

Defra (~2005) Inside Defra: How Defra works? http://www.larpnet.com/downloads/insidedefra.pdf

Defra (2011). Government Review of Waste Policy in England 2011. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/69401/pb13540-waste-policy-review110614.pdf

DeVellis, Robert F. (2017). Scale Development: Theory and Applications. Fourth Edition. Sage: Los Angeles.

Fagan-Watson, B., Watson, T., Elliott, B., Vanner, R., Shaw, B. and Morris, S. (2015). 'What range of methods is available to evaluate the implementation and impact of complex environmental policies on complex systems? A report to the Department for Environment, Food and Rural Affairs. Policy Studies Institute. Defra, London.

Hakim, Catherine (2000). Research Design: Successful designs for social and economic research. Second Edition. Routledge: London and New York.

HM Treasury (April 2011). The Magenta Book: Guidance for Evaluation. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf

Nickerson, Raymond S. (1998). 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises.' Review of General Psychology, Vol. 2 No. 2, pp 175-220.

Serco and Eunomia (2014) Investigating the Impact of Recycling Incentive Schemes: Full report. https://www.serco.com/Images/Serco%20Eunomia%20Incentives%20Full%20Report_tcm3-44276.pdf

WRAP (2015). Analysis of recycling performance and waste arisings in the UK 2012/13. http://www.wrap.org.uk/content/factors-influencing-recycling-performance

Centre for the Evaluation of Complexity across the Nexus

University of Surrey
GU2 7XH Guildford, UK

Email: cecan@surrey.ac.uk
Website: www.cecan.ac.uk