Learning lessons for evaluating complexity at the nexus:

A meta-evaluation of CEP projects

Collingwood Environmental Planning

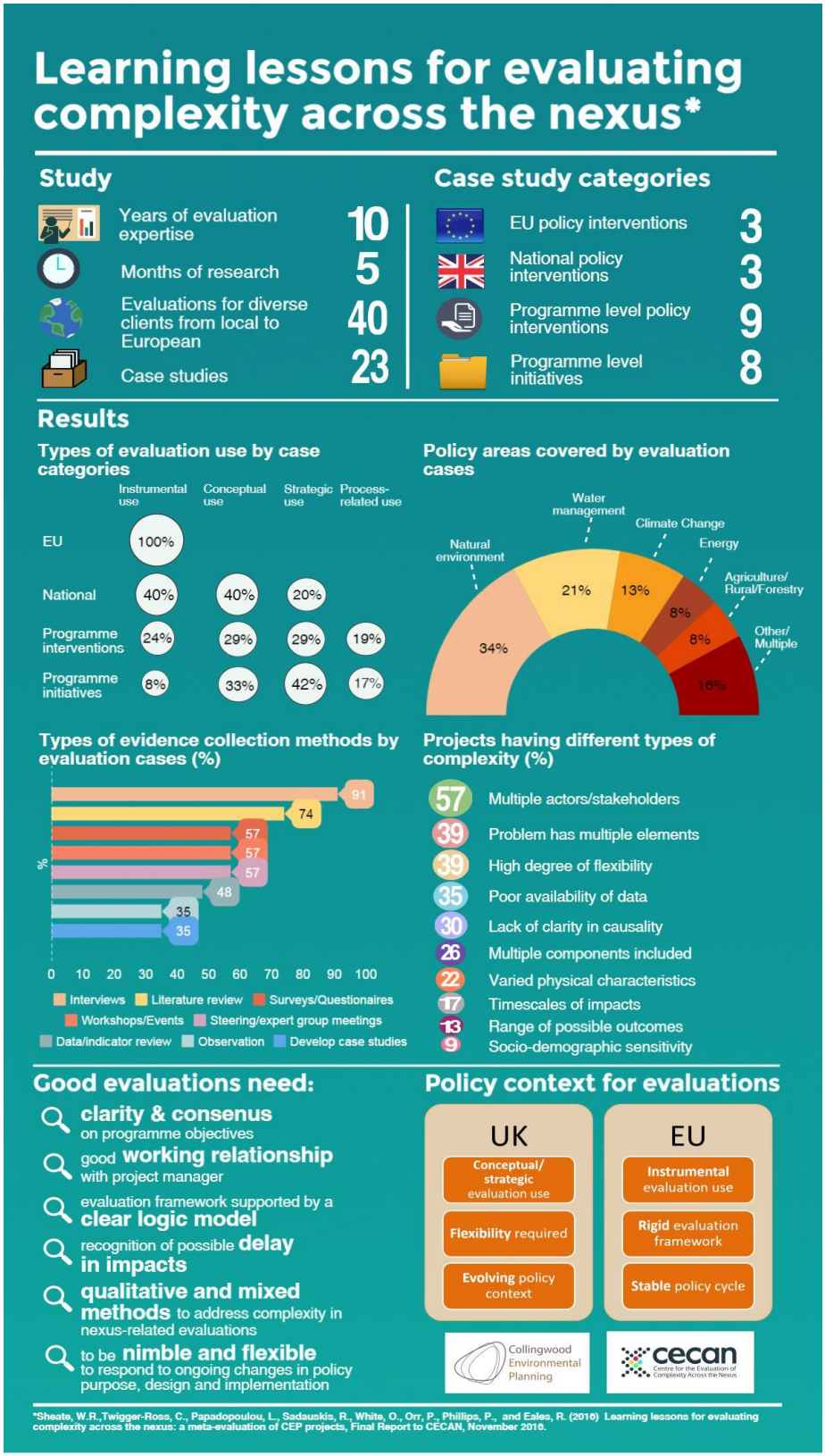
cecan
Centre for the Evaluation of Complexity Across the Nexus

Collingwood Environmental Planning Ltd

Project title:	**Learning lessons for evaluating complexity at the nexus: a meta-evaluation of CEP projects**
Project context	This project was carried out as part of CEP's contribution to the Centre for Evaluation of Complexity at the Nexus (CECAN). It was overseen by a project board of Dr Emma Uprichard, Dr Karen Scott and Dr Barbara Befani from CECAN
Project Team:	Dr William Sheate, Dr Clare Twigger-Ross, Liza Papadopoulou, Rolands Sadauskis, Owen White, Paula Orr and Ric Eales
Report details:	**Report title: Final Report** Date issued: 30 November 2016 Version no.: FINAL Authors: Dr William Sheate, Dr Clare Twigger-Ross, Liza Papadopoulou, Rolands Sadauskis, Owen White, Paula Orr, Peter Phillips and Ric Eales
Citation:	Sheate, W.R.,Twigger-Ross, C., Papadopoulou, L., Sadauskis, R., White, O., Orr, P., Phillips, P., and Eales, R. (2016) *Learning lessons for evaluating complexity at the nexus: a meta-evaluation of CEP projects*, Final Report to CECAN, November 2016.
Acknowledgements	This research was undertaken as part of CEP's involvement with CECAN, *The Centre for Evaluation of Complexity Across the Nexus*, funded by ESRC, NERC, Defra, BEIS, Environment Agency and the Food Standards Agency.

Contents

Summary



1. Introduction

CEP over the last 10 years (2006-2016) has undertaken an extensive range of evaluations in the natural environment arena, e.g. for Defra, Environment Agency, Natural England, Research Councils, Scottish Government, European Commission, Natural Resources Wales, Cefas, OECD, etc. These have all been published as publicly available documents, but core knowledge and learning on the delivery of these projects still resides within CEP staff. We have extensive experience of applying Magenta Book principles, among other approaches, and have managed and been involved in these projects over this significant timescale, and at a time when evaluation has risen up the political agenda as part of increasing accountability and value for money.

There is therefore a wealth of institutional memory to be mined from these projects on evaluation approaches and methods, such as, for example, the rationale for the choice of methods used, the challenges in undertaking complex evaluations in these areas, including the objectives of policy/policy interventions and the objectives set for evaluations, process *versus* outcome/impact focused evaluations, lessons learned, and barriers and enablers to the evaluation of complex policy/policy interventions and so on. In addition, many of the evaluations CEP has undertaken have included formative (process) evaluation as well as ex post (summative) evaluation of outcomes/impacts, often with explicit objectives of providing learning to the organisations/partners involved in policy implementation and policy interventions. We have also undertaken specific projects in relation to the counterfactual (a particularly challenging aspect in the field of complex policy), many have made use of logic models and/or theory of change and we have undertaken evaluations across the UK and devolved administrations, as well as in the European/international context.

A major gap in policy making is learning the lessons from past interventions and in integrating the lessons from evaluations that have been undertaken. This project – an important *meta-evaluation* of a sample of CEP evaluation projects (23 in total), was therefore undertaken as an intensive piece of research over 4-5 months (July-November 2016). The purpose of this meta-evaluation - to learn lessons from past policy evaluation around the NEXUS - fitted well within the 'Scoping needs' part of CECAN (and hence a reason for intensive activity). The outcomes of the meta-evaluation also provide some suggestions on evaluation practice for follow-up by CECAN and for researchers and practitioners.

1.1 Definitions

The Nexus

These are two definitions of the Nexus, the first from the CECAN website and the second from the Nexus shocks website:

*"'What works in practice' can be very difficult to ascertain, especially with policies that cut across the energy, environment and food Nexus domains, where urgent matters such as the 'energy trilemma', loss of biodiversity, climate change, poverty and challenges to health and well-being are entangled in complex ways." (*CECAN, 2016)

"In the past five years, there has been a surge of interest in the idea of the 'nexus', as a way of thinking about the interdependencies, tensions and trade-offs between food, water and energy security, in the wider context of environmental change.

It is widely understood that these different systems are inextricably linked. Efforts to improve sustainability in one domain without considering wider connections often prove inadequate. More integrated approaches are required, which move beyond sectoral, policy and disciplinary silos.

The nexus came to prominence in the water domain; an influential 2011 report from the World Economic Forum described water security as 'the gossamer that links together the web of food, energy, climate, economic growth and human security challenges.'

The concept gained further currency in the lead up to the Rio+20 Summit in 2012, and continues to attract the attention of a range of influential players in international science, policy, business and civil society" (The Nexus Network, 2016).

Both stress the interconnectedness of food-energy-environment domains and emphasise a need to consider that interconnectedness, complexity in order to understand how policies that cut across those domains work in practice. The first quote suggests that each of these areas are "Nexus domains" whereas the second quote emphasises the interconnectedness as the nexus.

The projects examined in this meta-evaluation focus on interconnectedness across nexus domains specifically: flooding, land use, climate change, catchment management, and biodiversity.

1.2 Aims

The aims of this project were threefold:

1. To learn the lessons from past policy evaluations;
2. To understand the factors that support or inhibit (barriers or enablers to) successful evaluations, where success is measured by
    a. Whether the evaluation meets its own objectives
    b. The impact that evaluation has - using four categories
        i. Instrumental – evidence has a direct impact on policy
        ii. Conceptual – evidence influences how stakeholders think about a policy area/issue
        iii. Strategic – evidence used for accountability and defending/promoting policy
        iv. Process – improved working processes in some way;
3. To explore the value of different types of approaches and methods used for evaluating complexity

A logic model (Logic Model 1) was developed to express the approach to the project (Table 1.1)

Table 1.1: Logic Model 1 for CEP meta-evaluation

Context	Inputs	Outputs	Outcomes	Impacts
The issue addressed and the context in which it is located?	What is invested e.g. money, skills, people, activities?	What has been produced?	Short and medium term results	Long term outcomes
• To learn the lessons from past policy evaluations, specifically: • To understand the factors that support or inhibit successful evaluations where success is measured: a) Whether the evaluation meets its own objectives; b) The impact that evaluation has - using 4 categories: Instrumental; Conceptual; Strategic; Process. • To investigate the value of different types of approaches and methods used for evaluating complexity	• Focused literature review; • Review of evaluations; • CEP brainstorm; • Project board meetings; • Review of relevant literature; • Interviews with CEP project managers.	• Characterisation of evaluations; • Spreadsheet of analysis; • Template to enable the inclusion of future evaluations for comparison that identifies types of impact, complexity and methods used; • Report for CECAN; • Seminar for CECAN; • Note for external folk on key issues. • List of methods/tools/strategies used • Case examples	• Increased understanding of barriers and enablers of successful evaluations, where success is measured in: 1) Whether the evaluation meets its own objectives; 2) The impact the evaluation has – across the four categories. • Increased knowledge of these issues within CECAN; • Increased knowledge of these issues and methods by government policy analysts with others outside of CECAN specifically policy analysts.	• Improved evaluation in complex areas of the nexus across various categories of policy evaluation • Improved understanding among evaluation practitioners through early identification of factors that may cause complexity in an evaluation and strategies to manage the complexity

1.3 Meta-evaluation Questions

To achieve the aims, the following key Evaluation Questions provided the evaluation framework and were used to guide the meta-evaluation.

1. Were the evaluations fit for purpose, and was their purpose clear? What lessons can we learn about assessing the effectiveness of the policy interventions?

2. Has the framing of the evaluation been more or less useful for understanding complexity (e.g. logic model, objectives led)? For example, in theory based approaches how useful has theory of change been in understanding complexity where the impacts are long-term (e.g. for biodiversity)?

3. What methods have been used for dealing with aspects of complexity found within environmental policy, e.g. long term nature of impacts, interrelationship of social and physical systems? Which methods appear to have been most effective? Were some methods and techniques more suited to certain types of complexity?

4. What factors lead to an evaluation being more (or less) influential of policy changes / outcomes / evaluation use?

For each of the case evaluations a series of specific details were collated ('evaluation specific questions') to provide the evidence required to consider the meta-evaluation questions. The evaluation specific questions correspond to and seek to test a series of mini-hypotheses ('hypothesis being tested') related to evaluation process and methods, drawing on CEP's institutional knowledge of what has worked well and less well in historic evaluations (Table 1.2)

A logic model (Logic Model 2) was developed to express these relationships, which is presented in Table 1.3.

Table 1.2: Evaluation sub-questions and working hypotheses

Meta-evaluation question	Evaluation specific questions *Information pulled out from each evaluation case example*	Hypothesis being tested *What is the assumption behind the question: mini-hypotheses*
Were the evaluations fit for purpose, and was their purpose clear? What lessons can we learn about assessing the effectiveness of the policy interventions?	What were the objectives of the policy or interventions? Were they clear and appropriate? Was there consensus on the evaluation objectives?	Clarity and consensus on the objectives of the intervention support effective evaluation (e.g. due to consistent implementation in different projects under a programme).
	What were the objectives of the evaluation? Were they clear and appropriate? Was there consensus on the objectives of the intervention?	Clarity and appropriateness of as well as consensus on the objectives of an evaluation will support efficacy.
	What were the circumstances within which the evaluation took place? Intervention governance arrangements (e.g. national/local tensions etc.)? Stable or evolving policy context (e.g. changes in higher-level political priorities etc.)?	A stable intervention governance and / or policy context will facilitate more effective evaluation.
	Project management context (steering group) etc.? Large or small steering group. Project manager (quality of).	Effective and efficient project management and governance leads to easier and more effective evaluation.
What methods have been used for dealing with aspects of complexity found within environmental policy, e.g. long term nature of impacts, interrelationship of social and physical systems? Which methods appear to have been most effective? Were some methods and techniques more suited to certain types of complexity?	What methods were used by the evaluation? What method/s were predominantly used? Why were some methods used / not used?	Some methods (or mixes of methods) will be more suited to managing certain types of complexity.
How has the framing of the evaluation been more or less useful for understanding complexity (e.g. logic model, objectives led)? For example, theory based approaches (e.g. theory of change to look at outcomes given that long-term impacts won't be able to be evaluated for e.g. biodiversity	How was the evaluation 'framed' (e.g. logic model, theory of change)? Was the evaluation framework developed as part of the evaluation, or provided by the commissioning authority? Were stakeholders involved in agreeing the evaluation framework?	An evaluation framework developed for a specific evaluation and with input from policy / intervention stakeholders is more likely to reflect the 'reality' of implementation and be effective in use.
What factors lead to an evaluation being more (or less) influential of policy changes / outcomes / evaluation use?	What happened to the evaluation? How (if at all) did it inform policy? Description and explanation, e.g. look across previous questions (e.g. Q1c on how policy may have been highly mobile during and/or after the evaluation, may have changed as a result of external/other factors; To what extent if at all did an evaluation influence that change in policy?)	A 'good' evaluation will not necessarily influence policy changes or decisions, which will be dictated by diverse factors beyond the scope of an individual evaluation to influence (e.g. funding, political changes etc.).

Table 1.3: Logic Model 2 reflecting the relationships between evaluation questions and hypotheses

Context	Inputs	Outputs	Outcomes	Impacts
The issue addressed and the context in which it is located?	*What is invested e.g. money, skills, people, activities?*	*What has been produced?*	*Short and medium term results*	*Long term outcomes*
To understand the factors that support or inhibit successful evaluations where success is measured a) Whether the evaluation meets its own objectives b) The impact that evaluation has - using 4 categories: Instrumental; Conceptual; Strategic; Process.	• Clarity of objectives of intervention being evaluated • Clarity of objectives of evaluation • Clarity of evaluation framework and its organisational development • Stability of governance context • Effective project governance e.g. responsive communication between project manager and client project manager • Time given to management of steering groups to achieve consensus on objectives etc.	Evaluation reports Learning events	Evaluation meets its own objectives Evaluation achieves impact: Instrumental; Conceptual; Strategic; Process.	

The meta-evaluation had the following Tasks:

Task 1: Focused literature review (Chapter 2) and development of the characteristics by which to categorise the evaluations.

Task 2: Clarify the research design (Chapter 3). We used a multiple embedded case study approach which was appropriate for the data. The projects were classified into types of evaluations; approaches; methods; types of complexity (e.g. around the Nexus), and sampled accordingly. Classification of the short list by categories using an Excel spreadsheet to document. Characteristics were listed as present or absent.

Task 3: Establish and agree Research Questions / Criteria for evaluating the evaluations, and having finalised the research questions, fill in as far as possible answers to specific evaluation questions drawing on final reports and other evaluation documentation for each of the chosen 23 examples. CEP project managers to review and validate where needed.

Task 4: Carry out interviews with each CEP project manager to complete missing or incomplete details, for example related to governance and/or what happened to the evaluation / policy.

Task 5: Complete analysis / assessment across all evaluations using the meta-evaluation questions to provide a thematic frame. This was completed initially in the Excel worksheet.

Task 6: Use complex mapping for looking at and mapping linkages between issues that emerge across all four meta-evaluation questions, i.e. are there linkages between the goodness of fit of the evaluations (Q1), the methods used (Q2), the framing and complexity dimensions (Q3) and what happened to it (Q4).

Task 7: Reporting - drafting of the final report.

1.4 Structure of the report

The report has the following structure:

Section 2: Summary of relevant evaluation literature, guidance and practice

Section 3: Case study design and approach

Section 4: Results and analysis

Section 5: Discussion: answering the evaluation questions

Section 6: Conclusions and next steps

References

Appendices

2. Summary of relevant evaluation literature, guidance and practice

The aim of this literature review is to provide context for this research around approaches to evaluation, focusing on meta-evaluations and identifying issues of complexity within evaluations, including any indication of where this complexity emerges from and how it can be dealt with.

This section is structured in three key sub-sections exploring these aspects, by reviewing existing guidance, academic literature and practice. In doing so it informs the approach adopted in this 'meta-evaluation' and identifies knowledge gaps that we then attempt to answer.

2.1 Evaluation guidance

There are various sources offering guidance on evaluation approaches and methodologies for policies. Depending on the source of this literature the terminology, scope and explicit objectives of the evaluation may vary. Often literature that may be relevant to evaluation refers to an 'appraisal' or 'assessment' with the distinction not always explicit. While the focus of this section and research is on 'evaluation', valuable lessons emerging from other relevant literature are captured.

Evaluation definition and objectives

Evaluation is an integral part of a broad policy cycle that the Green Book (HM Treasury, 2003) formalises in the acronym ROAMEF: Rationale, Objectives, Appraisal, Monitoring, Evaluation and Feedback. While evaluation is similar in technique to appraisal, it uses historic (actual or estimated) rather than forecast data to assess the policy effectiveness and efficiency, and it takes place during and after policy implementation. A widespread definition of evaluation can be found in the Green Book:

> *"Evaluation examines the outturn of a policy, programme or project against what was expected, and is designed to ensure that the lessons learned are fed back into the decision-making process. This ensures government action is continually refined to reflect what best achieves objectives and promotes the public interest."* (HM Treasury, 2003)

In a more simplified definition the Magenta Book (HM Treasury, 2011) describes evaluation as "*an objective process of understanding how a policy or other intervention was implemented, what effects it had, for whom, how and why.*" In practice, these questions and their responses are much more complex including considerations of how different features of the policy affected the way it performed and delivered, and how its outcomes varied across those it impacted upon: what worked for whom in what circumstances (HM Treasury, 2011).

The overarching objective of evaluation is to offer an unbiased assessment of a policy's performance by measuring outcomes and impacts in order to assess whether the anticipated benefits of a policy have been realised. A good evaluation, however, does not stop there but ensures that lessons are learned and communicated so that they may inform future proposals and policies. It therefore provides information on what could be improved in the design and delivery of a policy. In doing so it often involves an evaluation of the process of policy implementation as one of the factors influencing success.

Evaluation purpose and benefits

The purpose of the evaluation depends on 'what' is being evaluated and 'when' or in which stage of the policy cycle or the policy design and implementation process it is being carried out. An evaluation

that takes place alongside the policy's implementation can support the delivery of the policy by identifying what works well or less well and why. In doing so, it can help improve the effectiveness of the policy in meeting its objectives, while it also offers the opportunity for course correction if necessary. An evaluation taking place following the policy delivery can allow lessons to emerge that will inform the development of new policies.

Evaluation can have benefits for policy makers, practitioners and the public alike. Overall evaluations generate valuable, reliable evidence that can improve the effectiveness of policies and support the reinvestment of often limited resources (HM Treasury, 2011). Further, by assessing policy performance and effectiveness evaluations support government accountability demonstrating how public resources are spent and assessing the benefits realised. A number of sources reviewing evaluation in different contexts agree that by assessing the effectiveness of a project programme or policy, evaluation helps:

- clarify the objectives of the policy and identify criteria for success;
- improve programme and project management towards meeting those objectives;
- improve transparency and accountability for public fund spending by reporting on activities and assessing outcomes and impacts;
- improving future practice and policy by identifying what work well and less well and linking approaches to different contexts and impacts.

(Warburton et al., 2010)

Depending on when the evaluation takes place some of these benefits will occur for the project or policy they are designed to evaluate (ongoing evaluation), while others will only emerge retrospectively to offer valuable lessons for the future (ex-post evaluation).

Integrating evaluation into decision-making and practice

It is worth mentioning that a review of existing literature reveals that evaluation is a relatively new element in the field of environmental policy. In contrast there are scientific areas, such as the medical sciences, where evaluation appears to be more developed and integrated in the design of projects and policies. The fact that many policies and interventions around the energy, environment and food nexus do not include evaluation as an integral part of their design is one of the main reasons that there is limited literature on evaluation practices in these fields, while guidance keeps emerging.

An increasing recognition of the benefits of evaluations along with increased requests for accountability of public spending has led to a surge of interest in rigorous program evaluation, over the past decade, for a wide range of public policy problems (The Oxford Handbook of Public Policy, 2014). The Green book encourages government departments and agencies to *"consider how appraisals and evaluations are integrated with decision making processes and governance structures*".

However, for the benefits of the evaluation to be realised, especially when referring to a summative or ex-post evaluation, it is important that lessons are learned and feed into decision-making. One of the problems identified is the need to translate evidence into a format that is relevant to the user whether that is a government department, private or not for profit organisations or evaluation practitioners. Similarly to the decision made on the appropriate approach in evaluation, dissemination and reporting of the findings also needs to match the audience that it is aimed at. A format that may be fit for the purposes of an assessment from a government department's point of view may not necessarily be the most appealing for evaluation practitioners or those involved in the project or policy delivery. The integration of that knowledge may therefore need to employ more than the dissemination of an

evaluation report to embed learnings. Common practices and suggestions include presentation of the findings to the commissioning body and wider stakeholders, workshops or training with practitioners, publicising the results through articles, newsletters and presentations at conferences or producing new guidance (Warburton et al., 2010).

Identifying a need to integrate evidence into decision making the UK government established the 'What Works' approach and a network of independent centres aimed to support local practitioners and commissioners to:

1. Undertake systematic assessment of evidence and produce clear and actionable synthesis of evidence
2. Translate the evidence in a common currency enabling comparisons of effectiveness
3. Maintain the focus on the needs and interest of the users
4. Publish and disseminate the findings in a digestible format
5. Identify gaps requiring capacity building or further research
6. Support practice by advising by ensuring that projects and interventions can be evaluated effectively

(HM Government, 2013)

Evaluation use

The term evaluation use or utilisation refers to the way(s) in which evaluations and their findings affect operations, decisions and outcomes (Caracelli & Preskill, 2000; Kirkhart, 2000; Patton, 1985; Weiss, 1981) and it is key in demonstrating an evaluation's success. Although the aims and objectives of an evaluation are a good indication of the evaluation use, there are a number of factors influencing the actual impact of an evaluation. According to Balthasar (2009), these can include:

- Institutional factors, such as the organisation triggering the evaluation
- Environmental factors, such as the evaluation culture
- Process-related factors, such as mechanisms in place for stakeholder engagement or policy implementation

These factors are revisited again in Section 2.3 with reference to the complexity they introduce to evaluations.

A key element encompassed in both institutional and environmental factors above, is the human element, referring mainly (but not solely) to the intended users of the evaluation. BetterEvaluation notes that "*the use of an evaluation often depends on how well the report meets the needs and learning gaps of the primary intended users"*[1]. However, there is also an element of how evaluation is perceived by users. Peck and Gozalski (2009), like others before them, note that evaluations are often seen by commissioning bodies and organisations as 'ideas for change' rather than concrete improvements to be implemented. Adopting this attitude towards evaluation has a significant impact on how the various recommendations are perceived and to what extent (or whether) these are taken on board. A recent attempt to provide a framework for evaluation use by Peck and Gozalski (2009) is presented in Figure 2.3.

[1] http://betterevaluation.org/plan/reportandsupportuse

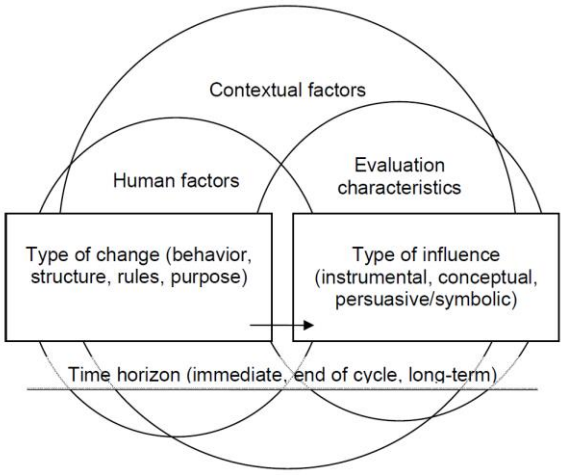


Figure 2.3: Integrated framework for evaluation use (taken from Peck and Gozalski, 2009)

Looking at theory and practice this model combines types of change (Downs, 1967 and Johnston, 1988 as in Peck and Gozalski, 2009) and types of influence/use (Kirkhart, 2000) in an integrated framework.

Bringing together the different perspectives emerging from theory and practice and adjusting them to fit the context of environmental policies, programmes and initiatives, this study adopts the following categorisation of evaluation uses (Table 2.1).

Table 2.1: Evaluation Use categories

Instrumental / Purpose-based use	• direct use of an evaluation's findings in decision making or problem solving • suggests changes to overall mission and aims
Conceptual use	• suggests changes in thinking or behaviours
Process-based / Structural use	• suggests changes on the basis of knowledge gained while undertaking the evaluation (Kirkhart, 2000) • suggested changes may refer to the organisation's or programme's structure
Strategic / Persuasive use	• evaluation is used to influence policy • can provide arguments in support of a political position (or not)

In line with earlier comments around attitudes towards evaluation, Peck and Gozalski's (2009) review found that very little instrumental use existed with most evaluation use being conceptual, as only a few organisations reviewed had implemented specific evaluation recommendations. Whilst, the role of conceptual use has been traditionally seen as less important, it has gained recognition in more recent literature (Peck and Gozalski (2009).

Focused evaluation guidance

The evaluation *'approach'* (i.e. the theoretical or philosophical perspective adopted in undertaking an evaluation) is often determined or influenced by the purpose of the evaluation and the 'worldview' of the evaluators, in a similar way that the evidence collection methods are determined by the nature of the project or policy to be evaluated. In a report for the Department for Constitutional Affairs, on evaluating public participation in central government, the authors note that quantitative methods are more widely used if the evaluation is designed for audit purposes, whereas an evaluation aimed at

learning will find qualitative methods, such as interviews and focus group discussions, more appropriate in understanding and describing what worked well and why (Warburton et al, 2010).

In addition to the **Green Book** (HM Treasury, 2003) and **Magenta Book** (HM Treasury, 2011) discussed earlier in this review, a selection of key guidance and frameworks have been identified and reviewed. These are presented in Table 2.2 followed by a column specifying the areas of policy where they have practical applications in and the key principles they identify as key to a successful evaluation. A more detailed description of each can be found in Appendix 4.

Table 2.2: Guiding principles for evaluation from key sources

Framework	Policy area of application / evaluation focus	Guiding principles for evaluation / Factors in assessing how successful the evaluation was
Qualitative Evaluation (HM Treasury 2012)	• A framework for assessing qualitative evaluations concerned with the development and implementation of social policy, programmes and practice. • A particular focus on the methods used in government-based evaluations	• **Contributory**: advancing wider knowledge or understanding about policy, practice, theory or a particular substantive field; • **Defensible**: providing a research strategy that can address the evaluative questions posed; • **Rigorous**: systematic and transparent collection, analysis and interpretation of qualitative data; • **Credible**: offering well-founded and plausible arguments about the significance of the evidence generated.
Quality in Public dialogue (Sciencewise, 2016)	A framework for assessing the quality of public dialogue.	**Clear scope for the evaluation:** • evaluation identified lessons emerging from the process and impacts • evaluation contributed to the design and delivery of the project • timing - the evaluation captured the entire project from the early stages and throughout the public engagement • appropriateness of the evaluation design in assessing success against objectives, participants' expectations, value for money, quality of engagement etc. **Analytical frameworks and criteria** • the specific evaluation framework used was identified • evaluation used clear criteria for the assessment of effectiveness • any assumptions and hypothesis were recognised and discussed in terms of how they affected the design and output of the evaluation • any unexpected outcomes and consequences identified and discussed **Evaluation reporting** • rationale for structure and form of the evaluation was discussed • the audience for the evaluation report identified • contributions of the evaluation to openness, transparency and accountability were discussed
EU Regulatory Fitness and performance Programme (REFIT) (European Commission,	Aimed at assessing whether the regulatory framework for a particular policy sector is 'fit for purpose'.	• provide an evidence-based critical analysis • establish whether the actions are proportionate to their objectives and delivering as expected

2015b)		
Better Regulation Guidelines on Impact Assessment (European Commission, 2015a)	Aimed at providing a framework for undertaking Impact assessments	• **Timing**: An Impact assessment is envisaged to happen as early in the process of policy development as possible. • **Content**: Includes clear descriptions of the problem, objectives, policy options, related issues and relevant impacts, providing an understanding of ex-ante uncertainties, possibilities and information sets
The Rainbow Framework[2] (BetterEvaluation, nd)	Acts as a checklist to ensure all factors have been considered in the evaluation design, implementation and dissemination.	• **Manage an evaluation:** understand and engage stakeholders; establish decision making processes; determine and secure resources; define ethical and quality evaluation standards; develop evaluation capacity • **Define what is to be evaluated:** develop initial description; develop programme theory/logic model; identify potential unintended results • **Frame the boundaries for an evaluation:** identify primary intended users; decide purpose; specify key evaluation questions; determine what 'success' looks like • **Describe activities, outcomes, impacts and context:** use measures, indicators or metrics; combine qualitative and quantitative data • **Understand causes of outcomes and impacts:** check the results support causal attribution; compare results to the counterfactual; investigate possible alternative explanations • **Synthesise data from one or more evaluations and generalise findings** • **Report & support use of findings in response to reporting requirements**, using **appropriate** reporting media, ensuring accessibility, developing recommendations and supporting use

Existing guidance suggests that the earlier an evaluation is planned for the easier it is to establish it as an integral part of the project and ensure that practitioners and stakeholders with responsibilities in monitoring and collecting data that will inform the evaluation are familiar with the objectives of the evaluation and the evaluation *framework* (or process).

Across documents reviewed above there is an emerging consensus on the key steps in the evaluation process. These can be summarised to include:

1. **Clarify 'who' will be involved**
2. **Clarify 'what' is being evaluated**
3. **Describe 'what' a successful evaluation looks like**
4. **Describe the evaluation activities and methods used**
5. **Understand causes of outcomes and impacts and compare results to counterfactual**
6. **Synthesise evidence and draw key findings**
7. **Present the results and recommendations**
8. **Disseminate and support use of the findings and recommendations**

[2] http://betterevaluation.org/plan

Evaluation types

The literature revealed a plethora of evaluation types. A range of criteria can be used for the distinction of different types of evaluations, but the traditional distinction emerges from the purpose of the evaluation which dictates whether an evaluation is formative or summative (see Table 2.3 for definitions). However, as an evolving field, evaluation literature and practice has gone beyond this simplistic characterisation to include a much wider range of evaluation approaches and types depending on what question the evaluation seeks to answer. Terminology is often confused among the literature and we simply have to accept that different people will use different terms to mean the same thing, e.g. types, approaches, even methods.

A comprehensive list of types can be found on the BetterEvaluation website. A detailed presentation of the range of approaches is beyond the scope of this focused literature review. This review therefore discusses complexity in evaluation prior to identifying and discussing which of the evaluation approaches seem to be suitable in the context of complex policy evaluations across the food, energy, and environment nexus. A more detailed description and characterisation of the different typologies of evaluation are presented in Appendix 1.
This review has found the following criteria (Table 2.3) have been used frequently, giving rise to different characterisations[3].

Table 2.3: Criteria and categories of evaluation

Criteria	Evaluation categorisations/ characterisations	Description of evaluation and its uses
Theoretical underpinning / Philosophy[3]	**Theory-based**	Also known as programme theory, theory- based evaluations are based on a theory of change or programme logic of "how an intervention, such as a project, a program, a strategy, an initiative, or a policy, contributes to a chain of events and finally to the intended or observed outcomes'' (Funnell and Rogers, 2011). They are usually visualised as a logic model often in the form of a linear sequence of inputs, activities, short-term and long-term outcomes and impacts. (Hagreaves and Podems, 2012)
	Realist	Realist evaluations have their methodological underpinning in the realism school of philosophy with applications in a range of disciplines. The realist approach assumes that there is no perfect approach that works for everyone, everywhere, anytime; rather the context really is crucial to a project or programme's outcomes. Hence, realist evaluations seek to establish causal relationships and develop an understanding of how an intervention works (or not) and why in a particular context and setting (Pawson et al, 2005). Realist evaluations often rely on logic models or theories of change as a way of looking at the entire system of an intervention from inputs to outputs and impacts and are therefore also commonly characterised as Theory-based. ***Also see relevant row under Synthesis method***

[3] Please note this list is not by any means an exhaustive list, but simply aims to demonstrate the range of characterisations that can exist illustrating their origins and most of all highlighting the overlapping nature of the various criteria/dimensions on the basis of which evaluations are categorised

	Idealist	Seeking a single "correct" hypothesis / answer, an evaluation that has its roots in the idealist philosophy is interested in generating (as opposed to testing) theory. It does not seek evidence to inform specific decisions but is rather "seeking concepts to provide enlightenment through new ways of understanding." (Gough et al, 2012)
Nature of the policy / intervention evaluated	**Experimental**	Experimental and Quasi-experimental evaluations are referred in the Magenta Book as 'Research Designs' aiming to "manipulate the implementation of the policy, or to exploit features which it already possesses, in such a way that a counterfactual can be estimated." (HM Government, 2011) Used to evaluate interventions that are innovative, experimental or include a high degree of uncertainty regarding their outcomes. The use of control groups is fairly common in experimental evaluations where a group is selected to match as closely as possible the characteristics of the project population. The most common method is Randomised Controlled Trials where an intervention is randomly applied to a selection of eligible groups or areas. Whereas commonly used in health interventions it is less common in policy, as policy allocation is not random by nature.
	Quasi-experimental	Quasi-experimental evaluation designs differ in that they do not use explicit randomisation but address counterfactual in other ways. It frequently involves identifying an existing less-than-random comparison group and addressing the selection bias.
	Non-experimental	Refers to traditional evaluation making use of observational data and evidence collected through monitoring to evaluate the impacts of a policy, programme or project on a set population or area.
Aim of the evaluation	**Formative**	A Formative evaluation is undertaken from the beginning of the project under review, and feeds into the development of the project. For that reason it can sometimes also be characterised as an Ongoing or Developmental evaluation.
	Summative	A summative evaluation is by nature ex-post and follows implementation. It is undertaken at the end of the project under review, providing an overview of the entire process.
Timing	**Ex-ante**	An ex-ante evaluation is undertaken prior to a project's implementation. It aims to gather information and carry out analyses to help to ensure that the delivery of the programme/policy objectives will be successful and that reliable evaluation will be subsequently possible. It is a fundamental tool for effective management and a formal requirement of the European Commission for any new (or the renewal of an expiring) expenditure programme.
	Ex-post	Ex-post evaluation is tries to measure a policy's / programme's or intervention's effects and impacts following implementation. Thus it is Summative in nature.
	Ongoing	An ongoing evaluation is undertaken in parallel with the project / programme or project under review. It can often include feedback loops and allow for course correction during implementation / delivery. Ongoing evaluation is thus Formative in nature
Approach	**Developmental / Adaptive**	Developmental Evaluation (DE) is an evaluation approach that can assist social innovators develop social change initiatives in complex or uncertain environments. DE 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. (BetterEvaluation, nd) By nature, Developmental evaluations are ongoing.
	Participatory	Participatory evaluation is an approach that involves the stakeholders of a programme or policy in the evaluation process. This involvement can occur at any stage of the evaluation process, from the evaluation design to the data collection and analysis and the reporting of the study. (BetterEvaluation, nd)
'What' is being evaluated'	**Impact Evaluation**	Systematic identification of the effects – positive or negative, intended or not – of a policy, programme, project. Impact evaluations provide estimates of the magnitude of outcomes and impacts. They are commonly used in informing policy-makers on decisions regarding the future of a policy / project.
	Thematic evaluation	An evaluation focused on a cross-cutting theme, fund, sector, modality, or service. (UN ESCAP, 2010)
	Programme / Project evaluation	An evaluation that focuses on the achievement of the results outlined in the logical framework of a project, often within the context of a broader programme. Most often, project evaluations are planned when the project is developed. (UN ESCAP, 2010)
Analysis method	**Meta-evaluation**	Uses methods such as meta-analysis / meta-review / meta-ethnography to synthesise and review qualitative studies and data. The analysis of this data aims to develop mid-range theories that have a direct application for particular defined areas of practice. They can be both Formative and Summative (Stufflebeam and Shinkfield, 2007) while Hanssen et al (2008) also describe a concurrent meta-evaluation that differs in that there is "continuous involvement, attendance at data collection events and external verification of the evaluation data"
	Qualitative Comparative Analysis (QCA)	QCA is based on the premise that the same outcome can be achieved in different combinations of conditions. It uses Boolean logic to explain pathways to a particular outcome and through an elimination process isolates the specific conditions / variables that are associated with a specific outcome.
Synthesis method	**Realist evaluation**	A realist evaluation synthesises evidence from a diverse range of sources to develop an understanding of "what works for whom and under which circumstances" (Pawson and Tilley, 1997) .These evaluations seek to identify and unpack the underlying mechanisms (M) that explain 'how' the outcomes (O) were realised and whether the context (C) in which they materialised made the intervention successful or not. Realist evaluations are therefore about hypothesising and testing a set of such CMO configurations. (Pawson et al, 2005)

(Sources: HM Government, 2011; HM Treasury 2003; Warburton, Wilson and Rainbow, 2006; BetterEvaluation, nd; Mackay et al, 2009; DCA, 2010; Dixon-Woods et al, 2005; Hagreaves and Podems, 2012; Pawson and Tilley, 1997; Pawson et al, 2005; Gough et al, 2012; UN ESCAP, 2010; Stufflebeam and Shinkfield, 2007; Hanssen et al, 2008; Westhorp, 2014)

Thus responding to any one of the above questions (e.g. what is the approach followed by the evaluation?) will yield a different categorisation / characterisation for the same study / review. It is therefore clear these are not mutually exclusive but in fact a number of categories may be used to describe the same evaluation, only different aspects of it, all being equally valid but bringing in focus a different dimension / element of the evaluation.

Reviewing literature on evaluation that is more focused on social and environmental policies / interventions reveals a selection of strategies in evaluation are more common than others. Due to the different criteria used to categorise evaluations, the same evaluation may indeed be characterised by more than one category describing its different elements. This review of evaluation types has informed our categorisation of the projects examined for this review.

2.2 Meta-evaluations

Meta-evaluation refers to the retrospective (ex-post) assessment of one or more projects, policies or evaluations. The term meta-evaluation is not widely used in the UK but references can be found in the United States (US) literature going back several decades (Downe et al., 2012). The term was first used by Scriven (1969) who defined meta-evaluation as a *"second order evaluation".* He described it as the *"evaluation of evaluations"* and identified two forms – one that underlines the purpose of the evaluation by asking questions around the evaluation approach and methodology, and a second that is concerned with the robustness and credibility of the results. Stufflebeam and Shinkfield (2007) more recently also distinguished between proactive and retroactive meta-evaluation, noting the former is aimed at helping evaluation practitioners in designing and carrying out an evaluation, whilst the latter is designed to "*help audiences judge completed evaluations"*.

Note though that there is a third view of meta-evaluation emerging from Sanderson (2000) that offers a more holistic approach and is the one adopted in this study. Sanderson saw meta-evaluation as an overarching assessment that operates above the level of individual projects or policies or programmes and recognises the complexity and interconnectedness of policy agendas seeking to identify common patterns and understand interactions (Hanssen et al, 2015). This third view recognises that the two initial forms do not have to be mutually exclusive, but on the contrary benefits can emerge both in terms of validating results and providing insights and lessons for future evaluation design.

Adopting this definition this study seeks to retrospectively look across project, initiative, programme and national-level evaluations. Reviewing these evaluations will allows us to reflect on the evaluations' design, process, implementation and delivery, and to identify learnings and key messages emerging on 'when', 'how' and 'why' different approaches may be more suitable than others. Reviewing these elements will also allow us to understand how evaluation findings are perceived and what is likely to support their use (BetterEvaluation, nd).

Options in meta-evaluations

Table 2.4 presents options in undertaking meta-evaluations as identified by BetterEvaluation organisation (2013). A common element in these is the inclusion of a 'personal factor' through the engagement in the meta-evaluation of people who were involved in the original evaluations either as evaluators, intended users, beneficiaries or stakeholders.

Table 2.4: Approaches in meta-evaluation

Approach	Description	Useful for
Beneficiary exchange	Facilitating a discussion of the evaluation findings with the project beneficiaries seeking feedback on project implementation.	• Developing recommendations
Expert review for meta-evaluation	Involves experts (individually or as part of a panel) reviewing the evaluation drawing on their expertise and experience of the particular type of program or project.	• Describing evaluations

Individual or group critical reflection	Involves asking particular individual stakeholders for their independent feedback or facilitating a group stakeholder feedback session on the evaluation findings	• Developing recommendations
Peer review for meta-evaluation	Involves reviewing the evaluation by using peers from within or outside of the organisation.	• Managing evaluations • Developing evaluation capacity • Supporting evaluation use

Robustness in meta-evaluation

Since meta-evaluation is relatively new as an approach, criticism on well-established approaches to systematic reviews where it draws from, such as meta-analysis and meta-ethnography, that dominated the field of in earlier years (Gough et al., 2012), is also relevant to the elements these approaches have in common. Meta-analysis is a systematic approach to synthesising quantitative evidence across studies commonly used in health interventions, whereas meta-ethnography involves combining data from qualitative research, especially ethnographic data, by translating concepts and metaphors across frequently social studies. Both of these methods aim to answer questions similar to meta-evaluations, such as "Do these types of interventions work?" or "For whom, in what ways and under what circumstances do they work?" Other studies also exist that refer to these and similar approaches as 'meta-narrative' approaches to evidence synthesis (Wong et al, 2014) or 'meta-studies' (Dixon-Woods et al, 2005). The difference in these approaches lies in the type of evidence synthesised, the particular methods used and the type of question they are most suitable to answer.

Based on these commonalities between these methods, the following criticism may hold credence for the robustness of meta-evaluations alike:

- **Selection and publication bias** caused by the use of a biased subset of the evidence / studies (Carpenter et al., 2011 on meta-analysis)
- **Tenability of verifying original study assumptions** due to the time lags and/or insufficient information reported in the original studies (Kromrey et al., 2005 on meta-analysis)
- **Heterogeneity of research** complicating the interpretation and comparison across research studies as a result of "*differences in substantive problems, theoretical perspectives, research methods, and researchers' goals*". (Fern and Monroe, 1996 on meta-analysis)
- **Lack of established methods and consensus** in how these studies should be undertaken. Further research into methods employed is required (Campbell et al., 2011 on meta-ethnography; Gough et.al on systemic reviews on meta-analysis and meta-ethnography)
- **Danger of over-generalisation or inappropriate application** due to lack of established terminology, methods and typology of these reviews (Gough et al. 2012 on meta-analysis, meta-ethnography and emerging approaches to systematic reviews)

Depending on the chosen option for undertaking a meta-evaluation, different criteria of what is considered a 'robust' approach or 'credible' data may apply. For instance, in an Expert Review (seen as one of the options in Table 2.4), the choice of qualified experts, the data they will be provided with for their review and the timing of their involvement will be key in ensuring robustness of the meta-evaluation. Knowledge about the nature of the specific methods used and good practice in those will be key in assessing the robustness of the approach, while recognising and clarifying similarities and differences across the studies reviewed is a first step in response to the above limitations.

Finally, with specific reference to meta-evaluations, a criticism could also emerge regarding the source of the evaluation. An evaluation that happens in-house and has the form of a self-assessment from the practitioner's side may be open to criticism on the objectivity of the findings, compared to that of an external, independent evaluator assigned will the particular role of evaluating a project or policy. While a number of documents make this distinction discussing implications on robustness, Levitt et al (2010) comment on the suitability of each evaluator depending on the case. In particular they note that: *"In some cases independent expert evaluation of the impacts of a project (or series of projects) will be valuable. In others it may be important to obtain feedback from those who have been subject to audit, inspection and scrutiny."* This more 'flexible' approach about 'who' is most appropriate to evaluate may be especially important in complex environments where the impact of the evaluation may be different in different parts of the environment, thereby requiring different kinds of objectivity and expertise.

2.3 Evaluating complexity

Evaluation can employ a variety of analytical methods to gather and assess information, and the choice of methods employed in any particular instance will depend on a wide range of factors. Choosing the right technique is crucial in the evaluation succeeding its objectives. The Magenta Book (HM Treasury, 2011) identifies the factors to be taken into account when deciding what sort of evaluation is necessary and appropriate: nature of the policy, its objective scale, complexity, innovation, form of implementation and future direction; the objectives of the evaluation and the types of questions it would ideally answer; the timing of key policy decisions and the information on which they need to be based; the types of impacts which are expected, the timescales over which they might occur, and the availability of information and data relating to them and other aspects of the policy; and the time and resources available for the evaluation.

Available literature on evaluating complexity often refers to programme theory, otherwise known as 'theory of change', 'intervention logic' or use of 'logic models' as an approach that allows evaluators to develop a causal chain between the programme inputs, activities, outputs and intended and observed outcomes (Rogers, 2000; Donaldson, 2005; Barnes et al., 2003; Davies, 2004; Pawson, 2006; Sanderson, 2000 – as in Rogers, 2008). Before reviewing these however, it is useful to develop a better understanding of 'what' complexity is in the context of environmental policy and intervention.

Complexity has been defined in varied ways, however a distinction commonly referred to throughout literature, draws on Glouberman and Zimmerman's (2002) distinction between what is 'complicated' (multiple components) and what is 'complex' (uncertain and emergent). As Rogers (2008) notes, these concepts have been adopted by a number of authors (Davies, 2004; Martin, 2007; Rogers, 2008; Synder, 2013). The distinction, explained in an evaluation context, is clarified in the following:

A. **Complicated project/policy and/or evaluation theory**: Elements that are inherent to the project or policy design, including multiple components, multiple actors/stakeholders, multiple and diverse activities, multiple simultaneous and/or alternative causal strands.
B. **Complex evaluation/programme theory:** Complexity refers to recursive causality (with reinforcing loops), disproportionate relationships (where at critical levels, a small change can make a big difference – a 'tipping point') and emergent outcomes.

(Rogers, 2008)

A common analogy used to distinguish a 'complicated' and 'complex' process is whether or not it can be reversed if 'broken'. For example, a clock is considered 'complicated' because it is possible to take it apart and put it back to together again, but an egg is 'complex' because breaking it is irreversible.

The Green book (HM Treasury, 2003) identifies a range of considerations to be taken into account when designing and undertaking an evaluation which according to the above definitions can contribute to complexity in the evaluation;

- Clarity of set targets and objectives
- Quality of data collected for the evaluation
- Impacts extending in multiple policy areas
- Geographic spread of the impact
- Equality - spread of impact across different societal groups
- Integration of regional perspectives into the policy making process
- Consideration / capturing intangible costs and benefits - externalities
- Additionality, 'leakage' and 'deadweight' effects' - Net impact after making allowances for what would have happened in the absence of the intervention.
- 'Displacement' and 'substitution' impacts

Time is another factor adding complexity as it is often the case that at the time of the evaluation taking place a number of the expected benefits of the policy have not yet materialised. Therefore, a review of literature and practice reveals that evaluations tend to assess short-term outcomes of the policy but do not always manage to capture the policy impact, as a longer term outcome. Jaffe, Newell, and Stavins (2005) discuss the complexity in the evaluation of policies related to environmental pollution and diffusion of new technologies to mitigate climate change. They identify the time lag between the policy and the expected benefits as well as a challenge emerging from the changing conditions as key difficulties in measuring policy success. They also note that the output or effect is often intangible, which gives rise to concerns of supporters of public investment in such technologies that the benefits of the policy will be understated and therefore the political support undermined.

The recognition of such uncertainties in evaluation is crucial particularly where timeframes are long, and considerable uncertainties exist with respect to the policy environment, baseline conditions and changing future conditions.

Table 2.5 below summarises a long list of factors that were found to introduce complexity in the policy arena being evaluated. The identification of these factors in a project could be an early signal of what has the potential to be a complicated and/or complex evaluation. However, it must be noted that, simply the existence of any of these factors is not sufficient to seal an evaluation as complex, since other factors may coexist and interact to mitigate their result. These complexity criteria are presented in what can be described as three different aspects of complexity:

- The complexity of **problem/issue** that the intervention being evaluated is trying to address;

- The complexity of **policy response** being evaluated;
- The complexity of **impacts** of the intervention being evaluated.

Table 2.5: Complexity criteria (Sources: HM Treasury, 2003; HM Treasury, 2011; Jaffe, Newell, and Stavins, 2005).

Problem-related complexity
• Problem has multiple elements
• Variability in the physical / environmental characteristics of the area / location
• Geographic spread / scale of the problem
• Sensitivity to socio-demographic characteristics of the area / target population (variability within the sample could mean complexity for the evaluation but specificity of the sample could mean difficulty in the transferability of the findings)
• Level of unpredictability in the problem (e.g. 'Tipping points' - Sudden, unexpected changes due to a small change having a big impact & Technological advancements – improvements in abatement technology)
Policy/Response-related complexity
• Multiple components / elements included in the policy/programme/initiative
• Multiple agencies / actors / stakeholders involved or targeted by the policy (may include conflicting interests)
• Degree of flexibility or tailoring / changes in the policy during implementation
• Geographic spread/ scale of the policy response
• Competing / interacting policies (at a UK or EU level)
Impact-related complexity
• Multiple types / range of possible / expected outcomes and impacts
• Unexpected / unintended impacts (positive or negative)
• Interactions between components of a policy (Number of interactions and combined impact – some interact to reinforce the combined impact while other neutralise each other's impacts minimising the combined impact)
• Lack of clarity in the causality between actions and impacts (difficulty in attributing causality)
• Timescales over which impacts might occur
• Availability of information and monitoring data relating to impacts

Identifying the factors that cause complexity is the first step in mitigating them. In *Evaluating Complexity: Propositions for Improving Practice* (Preskill et al, 2014), authors having identified the nonlinearity of systemic change in the social sector, call for governments, organisations and NGOs to "move beyond traditional, mechanistic strategic models and to take more of an "emergent" approach that better aligns with the complex nature of problems one wishes to solve. Their propositions for evaluating complexity include:

- **Proposition 1:** Design and implement evaluations to be adaptive, flexible, and iterative.
- **Proposition 2:** Seek to understand and describe the whole system, including components and connections.
- **Proposition 3:** Support the capacity of the system to learn by strengthening feedback loops and improving access to information.
- **Proposition 4:** Pay particular attention to context and be responsive to changes as they occur.
- **Proposition 5:** Look for effective principles of practice in action, rather than assessing adherence to a predetermined set of activities.

- **Proposition 6:** Identify points of energy and influence, as well as ways in which momentum and power flow with the system.
- **Proposition 7:** Focus on the nature of relationships and interdependencies within the system.
- **Proposition 8:** Explain the non-linear and multi-directional relationships between the initiative and its intended and unintended outcomes.
- **Proposition 9:** Watch for patterns, both one-off and repeating, at different levels of the system.

Reviewing what could be described as an appropriate evaluation mix for evaluating complexity, one comes across references to realist evaluation (Magro and Wilson, 2013; Pawson, et al 2005; Pollitt, 2013; Woods et al., 2005). According to Pawson et al. (2005) a realist review or realist evaluation, offers a model of research synthesis which is designed to work with complex social interventions or programmes. Grounded in the theory that underlines a programme or intervention it seeks to collect evidence from a diverse range of available sources. The review combines theoretical understanding, empirical evidence, case studies and formal reports with qualitative data from interviews often undertaken with those involved in the evaluations, to explain the relationship between the context in which the intervention is applied, the mechanisms by which it works and the outcomes which are produced. It provides an explanatory analysis aimed of what works for whom, in what circumstances, in what respects and how.

Hargreaves and Podems (2012) also argue that theory-based approaches are more appropriate in dealing with complex interventions as they provide early feedback about what is working or not, and why, thus allowing early intervention and course correction. In situations where there is uncertainty regarding the approach of the programme / project and the expected outcomes, such evaluations – also characterised as developmental - can prove more appropriate compared to formative and summative evaluations (BetterEvaluation, nd). Patton (1985) further suggests that the increased involvement of key stakeholders in the project delivery, decision-making and monitoring and evaluation is a more pragmatic approach in complex evaluations. Benefits of using such an approach include the development of a sense of ownership for those stakeholders involved who are also able to contribute local knowledge and insights for the evaluation. However, Hargreaves and Podems (2012) warn of challenges in stakeholders expressing contrasting views while others warn of these evaluations being regarded less objective (BetterEvlauation, nd)

2.4 Policy making

While traditional assumptions in relation to evaluation (e.g. in the Magenta book) are that a policy cycle exists and that evaluation occurs as a specific stage in that cycle (see Figure 2.1), it is also widely recongised that such a policy cycle is in practice often very fuzzy. Evaluation needs to recognise the fuzziness of the policy process because it affects if, how and when evaluation might have any influence on policy. Research by the Institute for Government (Hallsworth, 2011) noted that increasingly, as a result of decentralisation and devolution of responsibilities in policy making from the centre, that:

- Policy formulation and implementation are not separate, but intrinsically linked;
- The potential outcomes of the policy itself may change significantly during implementation;
- Complexity in public service systems often means central government cannot directly control how these changes happen;

- The real world effects policies produce are often complex and unpredictable.

Figure 2.1: Traditional representation of the policy cycle (Source: Defra, 2011.)

This means that in the UK policy making no longer (if it ever did) necessarily follows a typical *policy cycle* (Figure 2.1) where evaluation is part of the cycle that leads to refinement of the policy. Instead, policy making and implementation are now seen as part of a more dynamic system (Figure 2.2), termed 'system stewardship' by Hallsworth (2011). This is a more flexible and adaptive model, but also means that evaluation now faces purpose, design and implementation simultaneously and needs to be responsive to changing circumstances. Both policy making and its consequences are more complex and so evaluation needs to be responsive to this complexity, as well as complexity *intrinsic* to the subject matter of environmental policy and the nexus. Conventional performance indicators are often poorly suited to this increasing complexity and uncertainty; more flexible participatory approaches are therefore needed in evaluation that can deal with multiple perspectives of different stakeholders (Hallsworth, 2011).

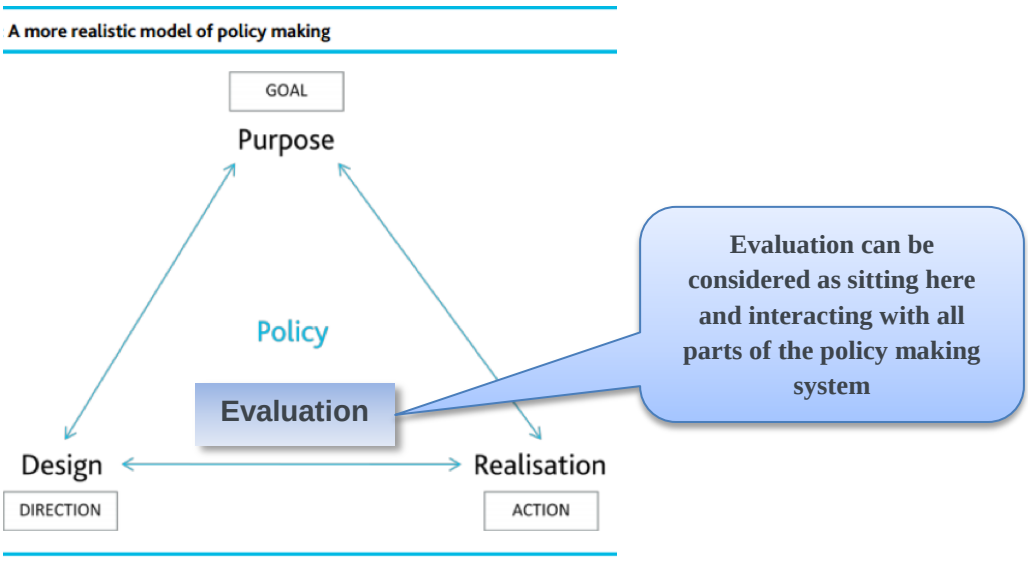


Figure 2.2: System stewardship model of policy making (Adapted from Hallsworth, 2011, Figure 5, p. 21)

3. Case study design and approach

A key aspect of the work was to characterise the evaluations across a number of key categories in order to provide a logical rationale for the case study design. An embedded case study design was considered to be appropriate given the nature and diversity of the *long list* of 43 past projects related to evaluation in some way undertaken by CEP in the last 10 years. The process of categorisation, using a master spreadsheet in Excel, was to facilitate selecting a sample of these cases that were internally consistent among themselves and could be used for meaningful comparison across the categories. This resulted in selecting **23** different projects (out of the original 43 long list) to focus on this meta-evaluation. In particular, it was found through this process that the defining characteristic that was most meaningful in terms of internal consistency was '**context**', rather than the type of evaluation, e.g. whether summative or formative, process or impact, not least because evaluations were often combinations of these).

A multiple embedded cases study design was created, selecting 23 cases (from the original long-list of 43) which could be broadly categorised as follows[4] (see):

- Policy interventions
  - EU policy interventions (EUP) (3)
  - National policy interventions (NP) (3)
- Programme level interventions/initiatives
  - Programme level *policy* interventions (PPI) (9)
  - Programme level *initiatives* (i.e. not linked directly to implementing specific policy) (PI) (8).

Cases eliminated by the selection process were those where:

1. CEP had had only a minor role, e.g. in data collection, or
2. the evaluations were very small scale, or
3. they might otherwise be in a category of one , or
4. there was no final report available.

Final reports were collated for each case as part of the categorisation process, to inform that categorisation and case study design and to complete a case study template for each case (see Appendix 3 for case study templates).

[4] The numbers in brackets indicate the number of cases reviewed under each category

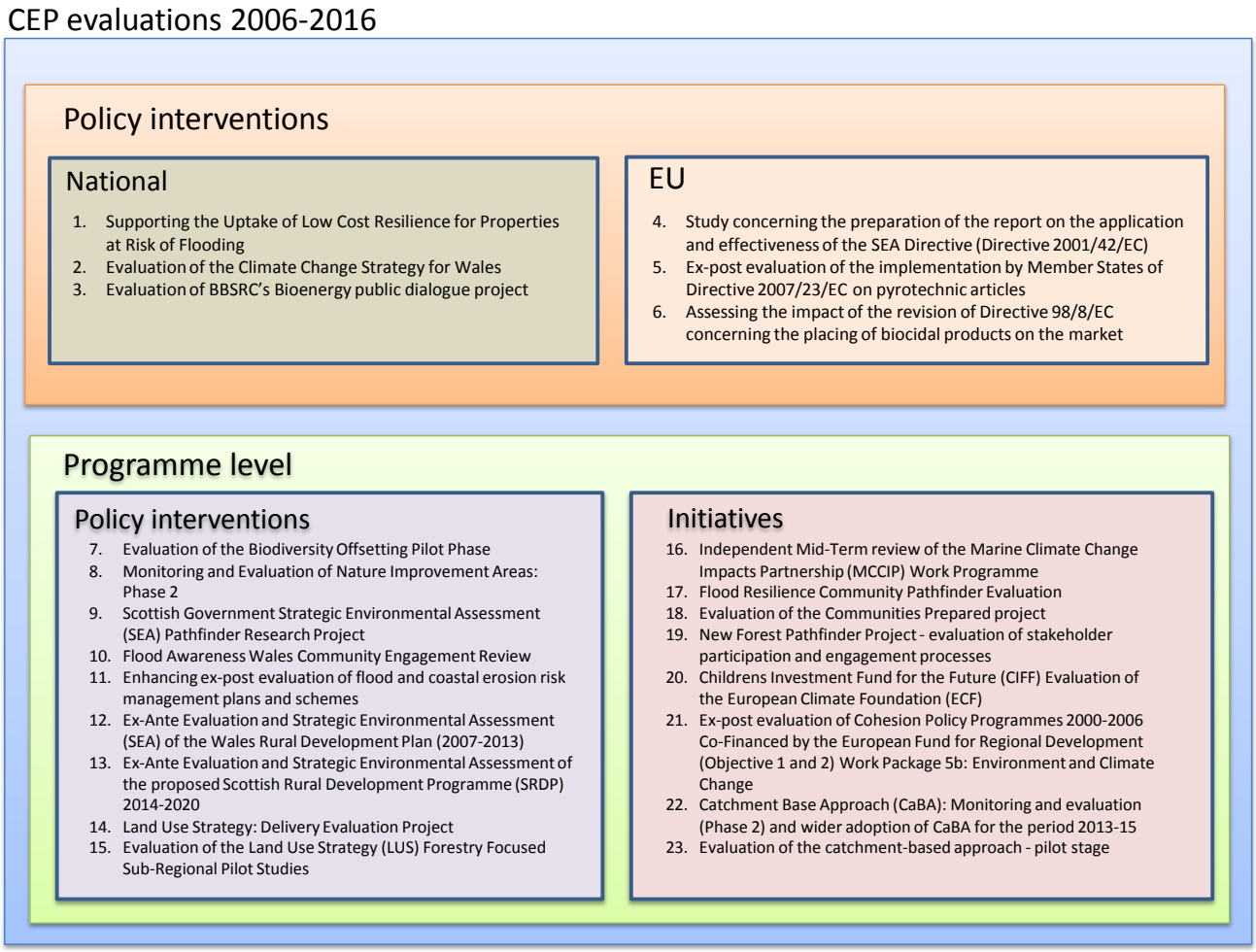


Figure 3.1: Complex policy evaluations around the NEXUS – Embedded case study design using 23 CEP projects

Alongside the development of the embedded case design, the categories for classification of the projects were developed. The categories chosen were those which related to the meta-evaluation questions, enabled comparison between the projects and came out of the literature review around complexity and evaluation. The projects chosen for the meta-evaluation were diverse and therefore categories were needed that could describe all the projects. The evaluation categories that were agreed upon are listed in Table 3.1 with more detailed definitions in Appendix 2 Tables A2.1-A2.7. These categories are then reflected in the case templates (Appendix 3) written up for each of the 23 cases.

Table 3.1: Evaluation categories

Category	Explanation
Scale	Geographical scale: local, regional, national and multi-national
Policy area	This describes the policy area that the evaluation is focused in, using UK government policy classifications: natural environment, agriculture/rural/forestry, energy, water management, climate change, other/multiple
Type of evaluation	This describes the evaluation approach used: *formative, summative, developmental, participatory, theory-based, ex-ante, ex-post, experimental, quasi-experimental, non-experimental.* - These are not mutually exclusive categories as some evaluations were combinations of different types.

Category	Explanation
Data collection methods	Methods used to collect data: *literature review, data/indicator review, observation, surveys/questionnaires, developing case studies, interviews, workshops/events, steering group/expert advice, participant diaries*
Types of complexity	Three areas of **complexity** were defined: • ***issue-related complexity:*** *problem has multiple elements, variability in the physical characteristics of the area, geographic scale of the problem, sensitivity to socio-demographic characteristics of the area, unpredictability in the problem* • ***policy/response-related complexity:*** *multiple components included in the policy/programme/initiative, multiple agencies/stakeholders involved or targeted by the policy, high degree of flexibility or tailoring/changes in the policy during implementation* • ***impact-related complexity:*** *multiple types/range of possible/expected outcomes and impacts, unexpected/unintended impacts (positive/negative), interactions between components of a policy, lack of clarity in the causality between actions and impacts (difficulty in attributing causality), long timescales over which impacts might occur, poor availability of information and monitoring data relating to impacts*
Evaluation use	Four types of **use** were examined: ***instrumental*** *– evidence has a direct impact on policy,* ***conceptual*** *– evidence influences how stakeholders think about a policy area/issue,* ***strategic*** *– evidence used for accountability and defending/promoting policy, and* ***process- related*** *uses – improved working processes in some way*
Budget	Six bands of budget were included in this category: *£20,000; £21,000 – £50,000, £51,000 - £99,999, £100,000- £199,999, £200,000-£300,000*

3.1 Method of classification

Once the categories had been agreed each of the projects was classified according to those categories by one member of the CEP research team using a ✓ to denote presence or absence of the category. For ease of use an Excel spreadsheet was developed listing all the projects and the categories. Once the first classification had been carried out each of the CEP project managers was asked to verify the classifications for their projects and to choose which three aspects of complexity were most reflected within their projects, and the types of evaluation use (where known). The classification process was iterative, with new categories being added through discussion with the project board and wider CECAN community. Specifically, the evaluation use category was divided into four sub-categories to capture the different types of use and the budget category was added to give an indication of the range of budgets associated with each evaluation.

3.2 Answering the specific evaluation questions

Once the classification process was finished, each of the projects was examined in relation to the specific evaluation questions with responses to those questions provided from project documentation e.g. final reports, supplemented by short interviews with the CEP project managers. All responses were recorded in the Excel spreadsheet with a clear distinction between sources being made (different coloured text). From these individual responses themes, messages, observations and examples were drawn out for each of the questions across all projects within a case category, e.g. EU policy interventions. These were recorded within the Excel spreadsheet. More work was carried out by two members of the team to draw out key themes which were then used to compare between the case categories and to produce the complex mapping of themes (see Section 4.2).

3.3 Thematic analysis and complex mapping approach

Our approach to thematic analysis is firmly rooted in qualitative research, drawing on grounded theory where key themes are identified and coded from the data.

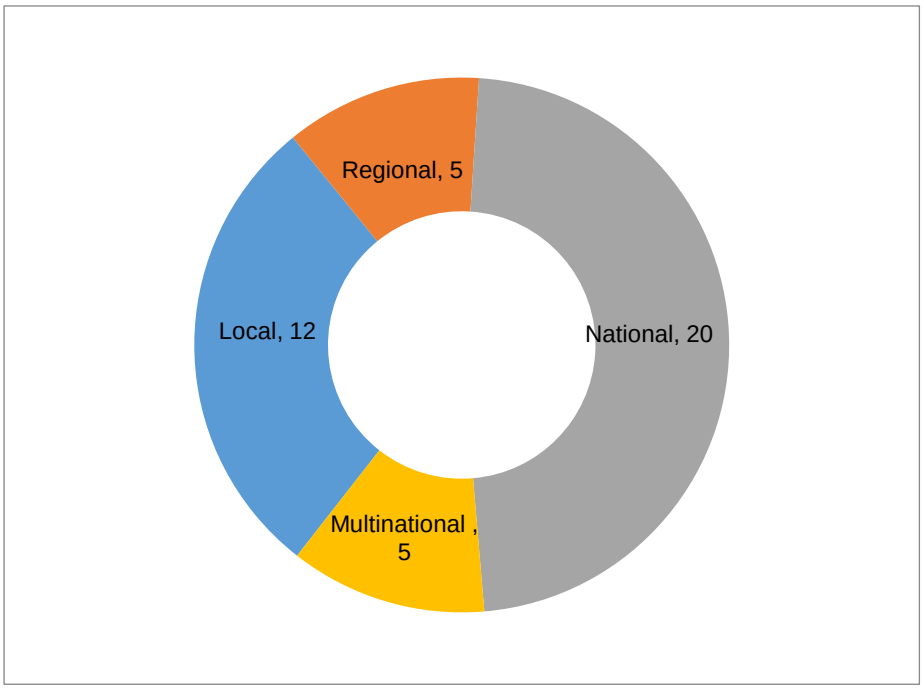
Emerging themes from the spreadsheet of cases were clustered and reviewed by two team members. Common themes across the case study categories were identified, some of which were pre-determined codes derived from the hypotheses/evaluation questions (e.g. clarity of objectives, stability of policy, existence of explicit theory of change); others emerged from the data, e.g. strength or weakness of the policy cycle. It was clear from the case study data that EU evaluations operated in a different policy context to the national and programme level interventions and initiatives.

The links between thematic issues (arrows, direction and effect – positive or negative) were mapped using the evidence from the data (final reports and interview data) and/or from the logical application of causal chain/network analysis and summarised in Table 4.1).

4. Results and analysis

4.1 Summary results

The characterisation exercise of all 23 selected CEP evaluation projects is summarised in Figure 4.1 - 4.14 below. Note that projects can be characterised as having multiple characteristics, which means that the total number can add to more than n (=23). The figures are ordered in the same sequence as Table 3.1: Scale, Policy area, Type of evaluation, Data collection methods, Types of complexity, Evaluation use, and Budget categories.



4.1: Evaluation scale

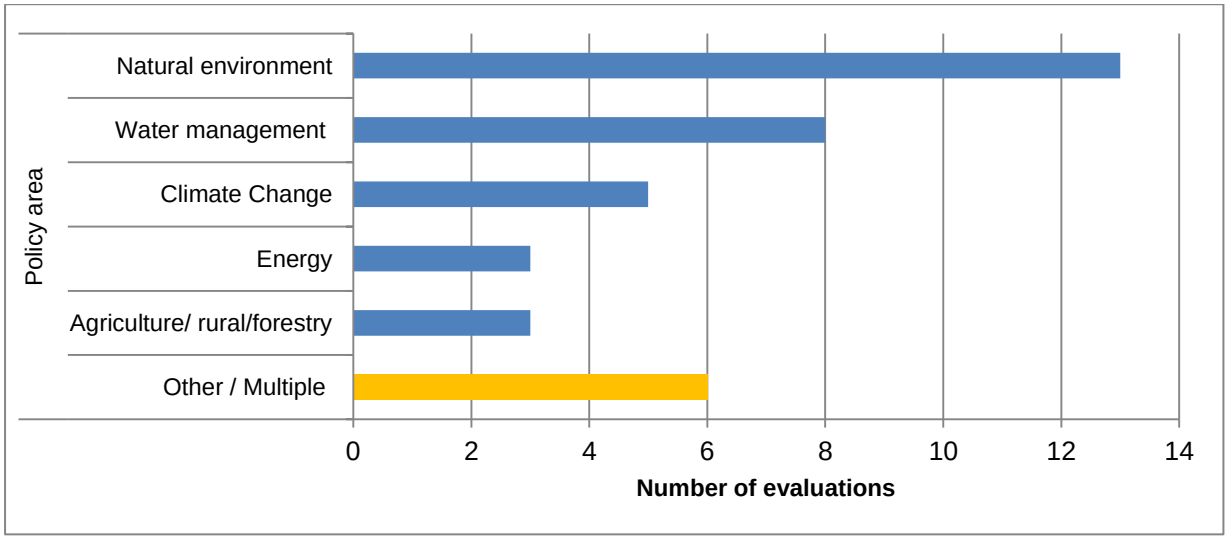


Figure 4.2: Policy area

Comments

- All policy areas are nexus issues
- Majority of projects are with the natural environment policy area
- Example nexus issues within that policy area include: land use, biodiversity offsetting, rural development, nature improvement areas, EU cohesion policy on environment

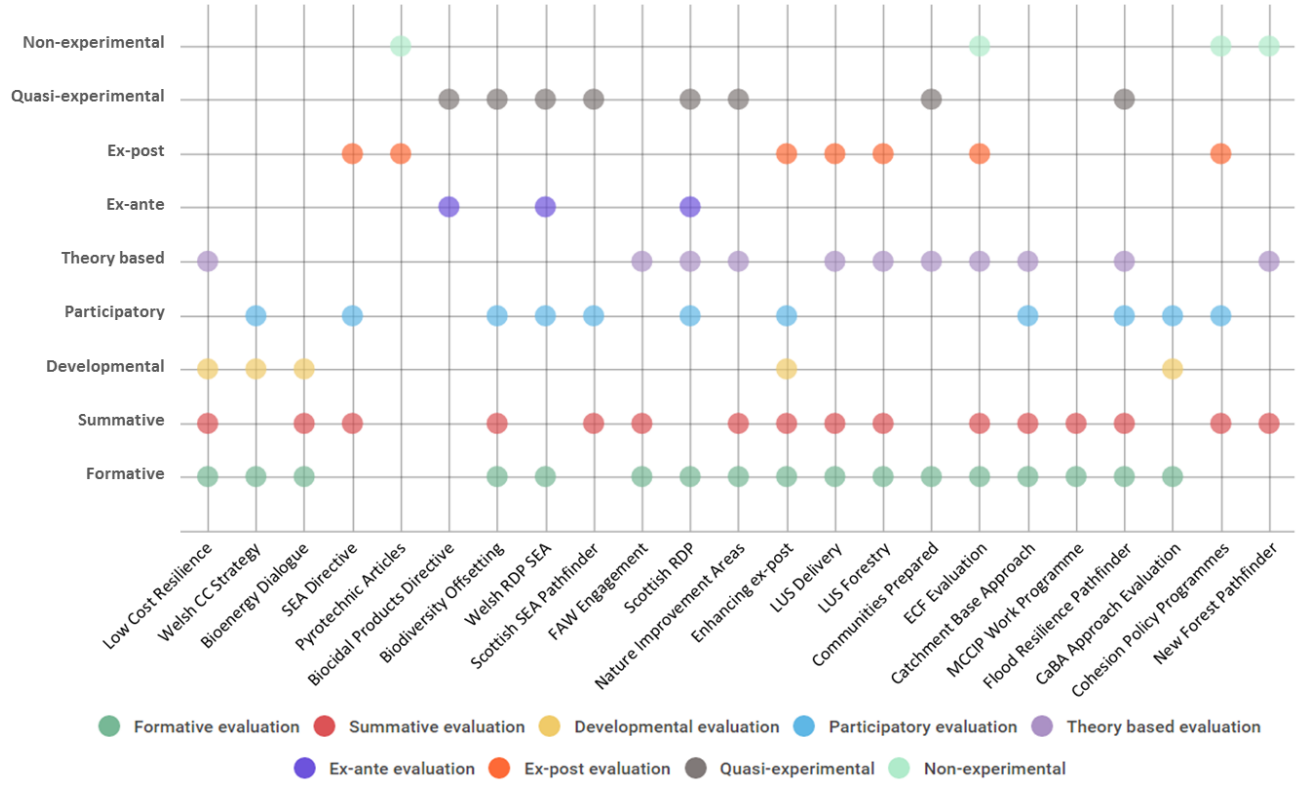


Figure 4.3: Types of evaluation by project

Note: This Figure refers to the following total numbers: Formative evaluation, 17; Summative evaluation, 16; Developmental evaluation, 5; Participatory evaluation, 11; Theory based evaluation, 11; Ex-ante evaluation, 3; Ex-post evaluation, 7; Quasi-experimental, 8; Non-experimental, 4

Comments

- Experimental evaluation is not included as there were none amongst the projects we looked at.
- There was no categorisation that was exclusive. In fact each project fitted into between 2 and 5 types of evaluation categories.
- The above figure uses cross-tabs across the evaluation types (see Table 4.1), showing that for all projects that were Formative, Summative etc. more categorisations also applied. For example, across all formative evaluations, 12 are also categorised as summative, 5 also categorised as developmental, 8 as participatory etc.

Table 4.1: Types of evaluation – cross-tab

	Formative	Summative	Developmental	Participatory	Theory based	Ex-ante	Ex-post	Quasi-experimental	Non-experimental
Formative (17)		12	5	8	10	1	3	4	1
Summative (16)	12		3	7	9	0	6	4	3
Developmental (5)	5	3		3	1	0	1	0	0
Participatory (11)	8	7	3		3	2	3	5	1
Theory based (11)	10	9	1	3		1	3	4	2
Ex-ante (3)	1	0	0	2	1		0	3	0
Ex-post (7)	3	6	1	3	3	0		0	3
Quasi-experimental (8)	4	4	0	5	4	3	0		0
Non-experimental (4)	1	3	0	1	2	0	3	0	

Comments

- Formative and Summative evaluations are commonly used in conjunction to satisfy project needs (12 projects satisfied both categorisations)
- There have been no ex-post evaluations that included an ex-ante evaluation of the project.
- All of the developmental evaluations were also formative (3 out of 3)
- Almost half of all participatory evaluations were quasi-experimental (5 out of 11)
- All of the ex-ante evaluations were quasi experimental

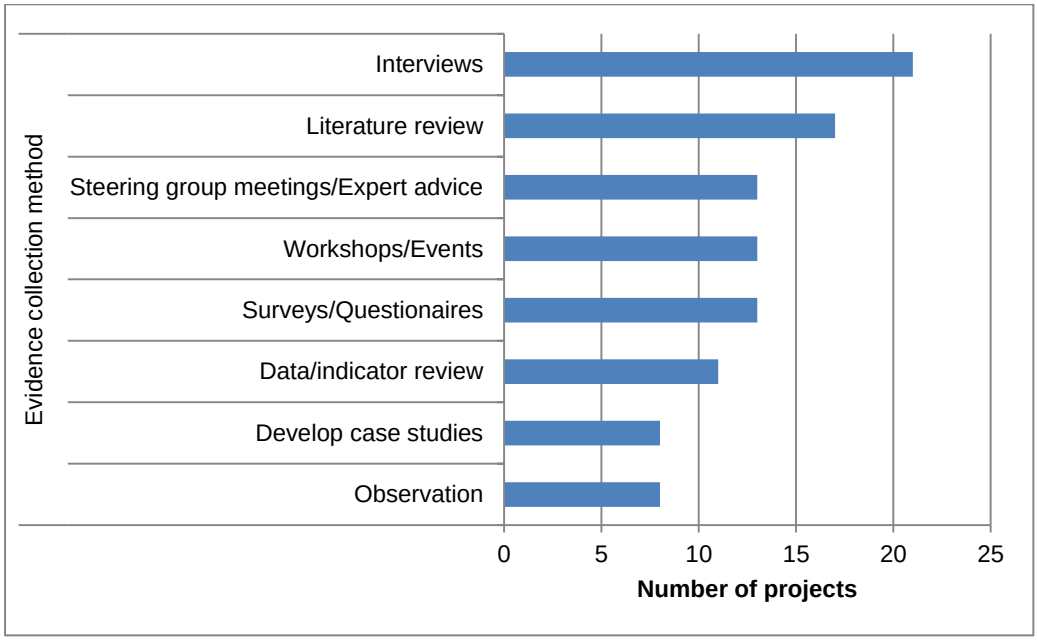


Figure 4.4: Types of evidence collection methods

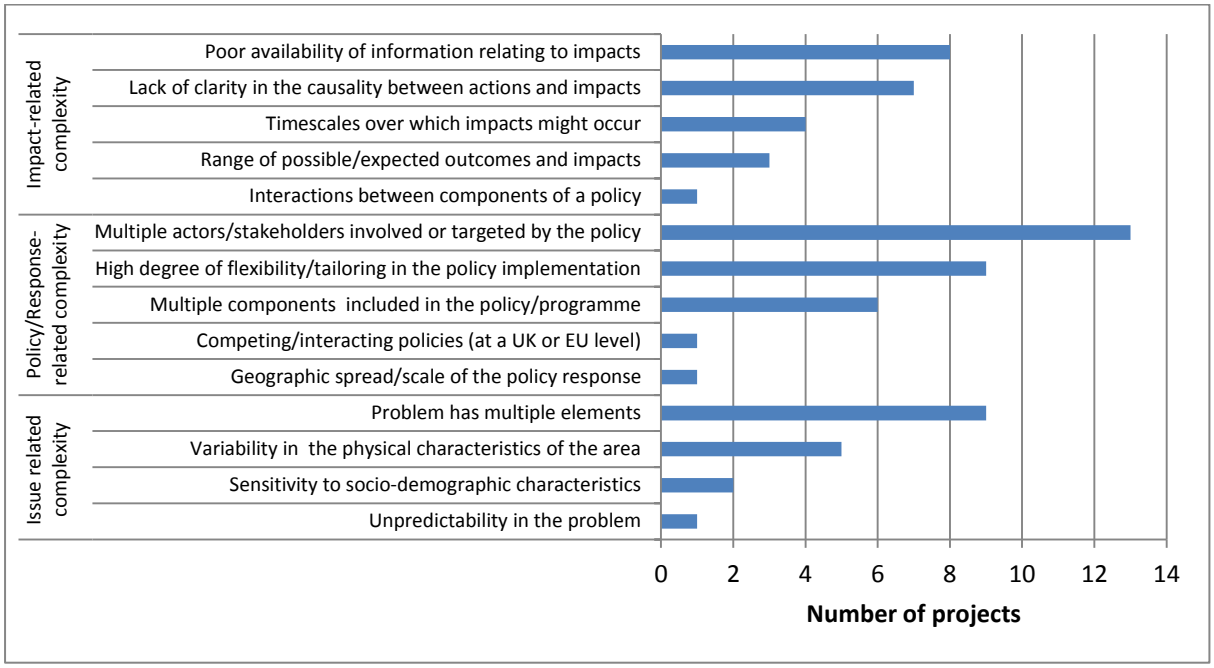


Figure 4.5: Types of complexity involved

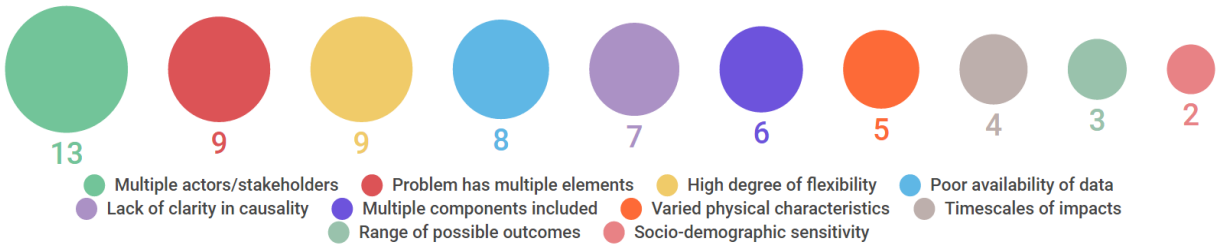


Figure 4.6: Number of complexity types identified across projects
Note: Only includes issues identified more than once

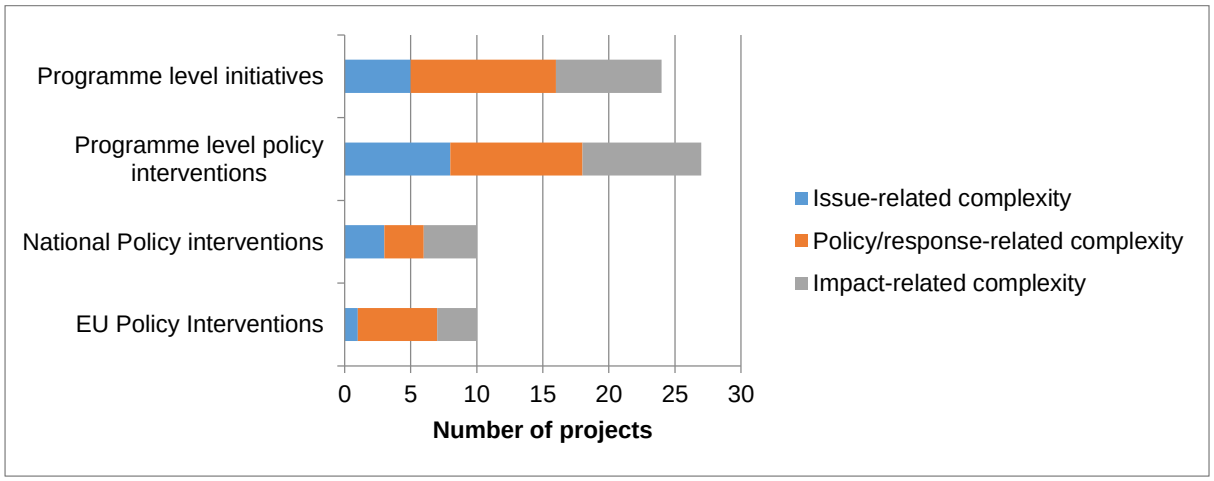


Figure 4.7: Types of complexity exhibited by case categories

Comments

- While some issues around complexity were faced by all types of projects, such as the availability of information and monitoring data, others tended to be specific to the nature of the project. **The following results need to be caveated on the small number of projects reviewed and included in the Figure above.** As such:
    - the majority of projects (5/7)dealing with a lack of clarity in the causality chain between actions and impacts were Programme level initiatives
    - the majority of projects (5/7)dealing with multiple components were also Programme level initiatives
- Policy/response related complexity is the main sourse of complexity existing in EU policy interventions, while it remains a considerable portion of identified complexity in Programme level initiatives.

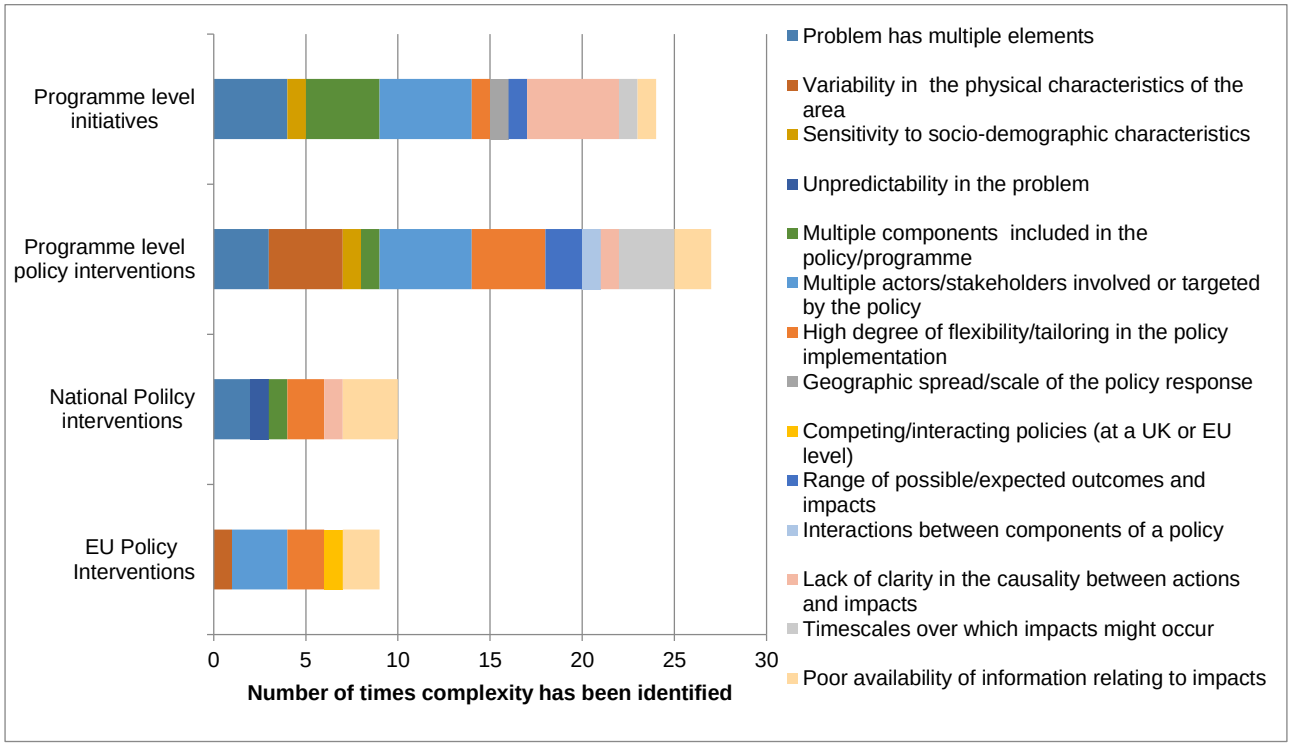


Figure 4.8: Further breakdown of types of complexity exhibited by case categories

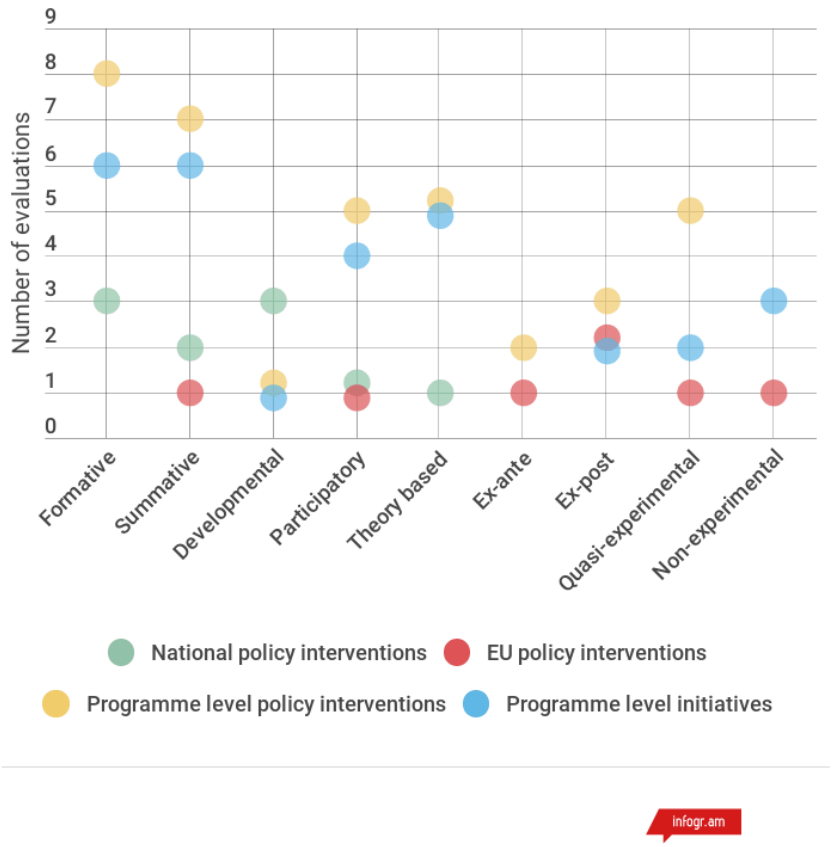


Figure 4.9: Number of evaluations by type and case category

Comments:

- Programme level policy interventions and programme level initiatives tend to be formative and summative, and often participatory.

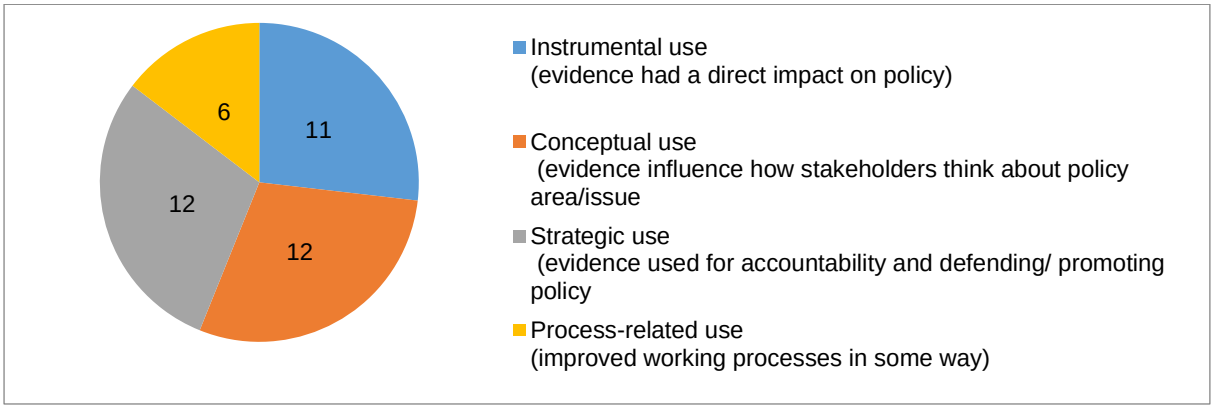


Figure 4.10: Types of evaluation use exhibited by CEP evaluation projects

Note: This Figure refers to a total number of 22 projects as one of the evaluations was ongoing at the time of authoring this report and the use of the findings remained to be seen.

Comments:

- Evidence from CEP evaluation projects of all types of evaluation use

- Many projects, where evaluation use/impact is known, tend to have conceptual or strategic use.

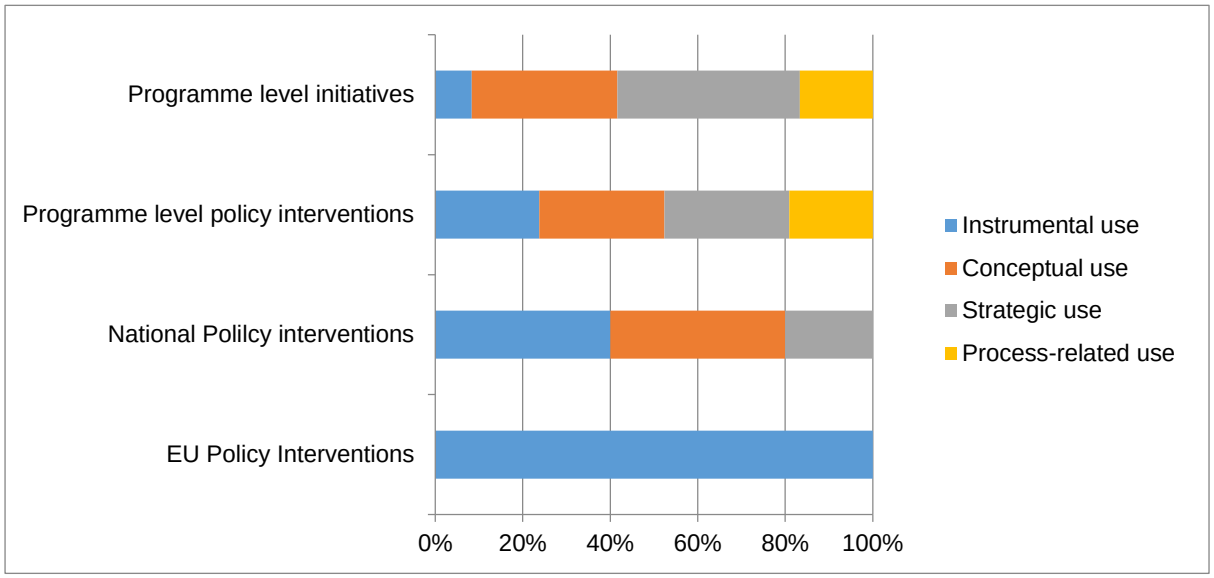


Figure 4.11: Types of evaluation *use* exhibited by case categories

Note: This Figure refers to a total number of 22 projects as one of the evaluations was ongoing at the time of authoring this report and the use of the findings remained to be seen.

Comments:

- All EU evaluations had instrumental use (as they are designed to)
- Instrumental use is less common at programme level policy interventions and an exception for programme level initiatives (as might be expected, given the lack of clear policy context)

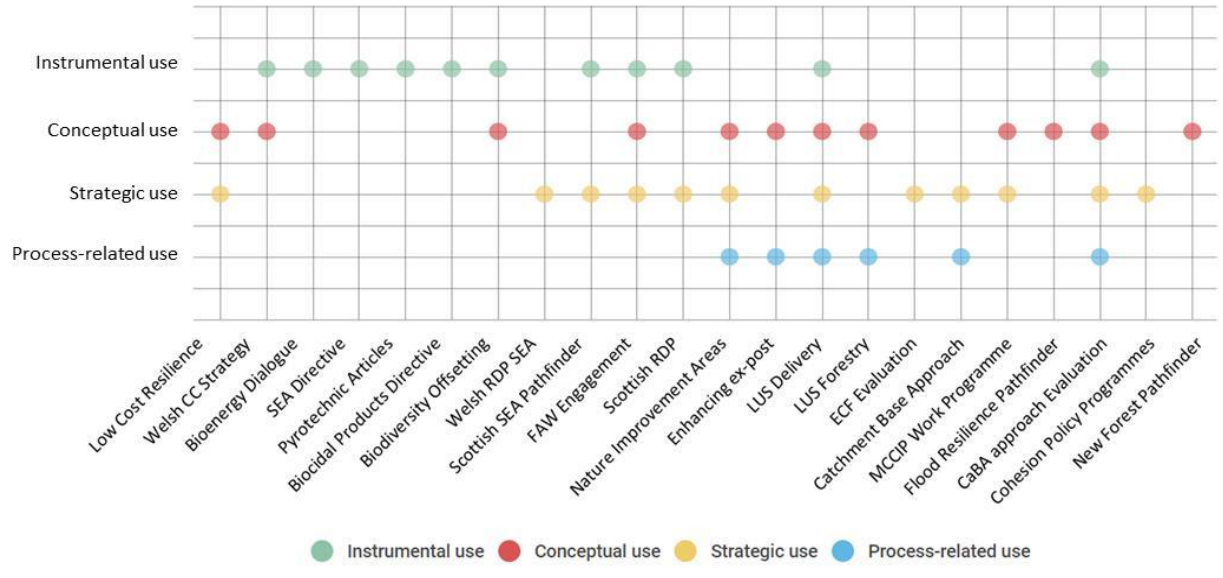


Figure 4.12: Types of evaluation use by project

Comments

- All of the EU policy intervention evaluations had instrumental use

- National Policy intervention evaluations (projects in the beginning of the figure) tend to have instrumental use. Equally none of them had a strategic or process-related use
- Programme level policies and initiatives were the only ones that had a process-related use.
- Only a very small number of projects (2 out of 23) covered all 4 types of evaluation use
- Almost half of the evaluations that had strategic use also had process related use (4/10) and almost all of those (3/4) also had conceptual use.

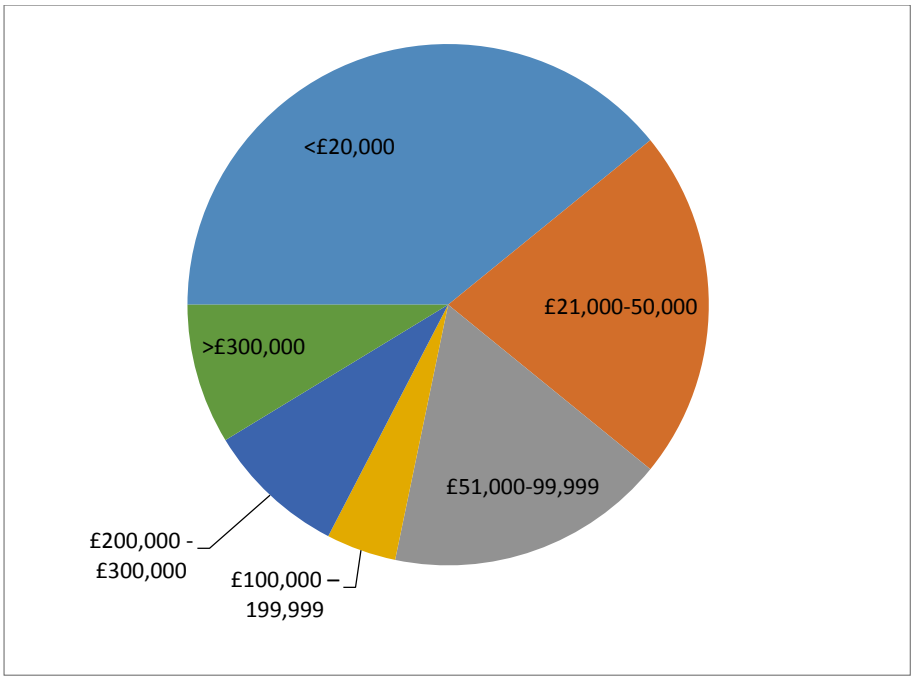


Figure 4.13: Evaluation projects by budget category

Comments:

- A large proportion (two thirds) of evaluation projects undertaken had budgets under £50,000.
- The largest single budget category was under £20,000; these smaller budgets tending also to reflect shorter timescales for evaluation and the more qualitative nature of the types of evaluations being undertaken.

4.2 Complex mapping of emerging themes

Table 4.1 below summarises the key themes that emerged from the evaluations, highlighting similarities and differences among the multiple embedded case studies. The emerging themes were used as the basis for complex mapping, as described in Chapter 3.3.

Two maps were produced – one for the UK national and programme level policy interventions and initiatives (Figure 4.15), since from the meta-evaluation they had many similarities in their characteristics, and one for EU level policy evaluations (Figure 4.16). While there were only three EU evaluations considered, compared to a larger number of UK and programme level evaluations, the EU cases were typical of the types of evaluations undertaken routinely for EU legislative instruments and policies.

Themes emerging from mapping

The first and strongest message that emerges from the mapping of the UK national and programme level interventions/initiatives compared to the EU interventions is the stark difference in the nature of policy making and the role of evaluation – the existence or not of a strong policy cycle and flexibility or rigidity in evaluation that flows from that.

As noted in the literature review, in the UK policy making no longer necessarily follows a traditional policy cycle, but is more 'dynamic' in the interaction among the goals, direction and implementation of policy (Hallsworth, 2011). There appears, at least at face value, to be strong evidence for this model from the meta-evaluation of CEP projects – the lack of a strong policy *cycle* is clearly evident, and the need for flexibility and adaptive capacity in the evaluations. While in principle this appears to conform to the 'system stewardship' model, in practice what was observed was often an *absence* of clear policy context, and policy interventions with a distinct lack of an explicit theory of change, both of which you might reasonably expect to see within the system stewardship model. This contrasts strongly to the very clear policy cycle that exists at the EU level which in turn establishes a much more rigid monitoring, review and evaluation process as part of revisiting policy/legislation on a regular cyclical basis (Figure 4.14 below).

In many of the UK cases an explicit theory of change was first elaborated as part of the evaluation process, after the policy intervention had been initiated and sometimes after it had been running for some considerable time (months or years). An important lesson from this for policy making more generally is the need for policy to be more explicit as to its objectives and intervention logic – what is it trying to achieve and how is it expected to achieve it? Especially in a dynamic system stewardship model where there is iteration among the purpose, design and implementation of policy, being clear about the purpose is essential and having a theory of change from inception means there is a theory that can be validated, modified and revised dynamically as evidence becomes available. A theory of change is needed – under system stewardship or a more traditional policy cycle – precisely so that there is clarity when objectives are modified or expectations change as the policy evolves.

The EU evaluations invariably have more rigid prescription regarding monitoring, indicators, and evaluation frameworks and evaluation questions because of the need for consistency and comparability across all EU Member States, including the use of mixed methods – especially the use of formalised regular reporting and quantitative indicators - and use of qualitative semi-structured interviews/focus groups with stakeholders. The evaluations invariably have instrumental use at the EU level because they are designed to do just that, in comparison to many UK evaluations which appear from this meta-evaluation (inasmuch as there is evidence that they are actually used) to have more strategic or conceptual use as they feed into in a more dynamic way to policy evolution. There is a much less obvious policy cycle and so the use of evaluation is less likely to be obviously instrumental in the way that it can be in a strong policy cycle context.

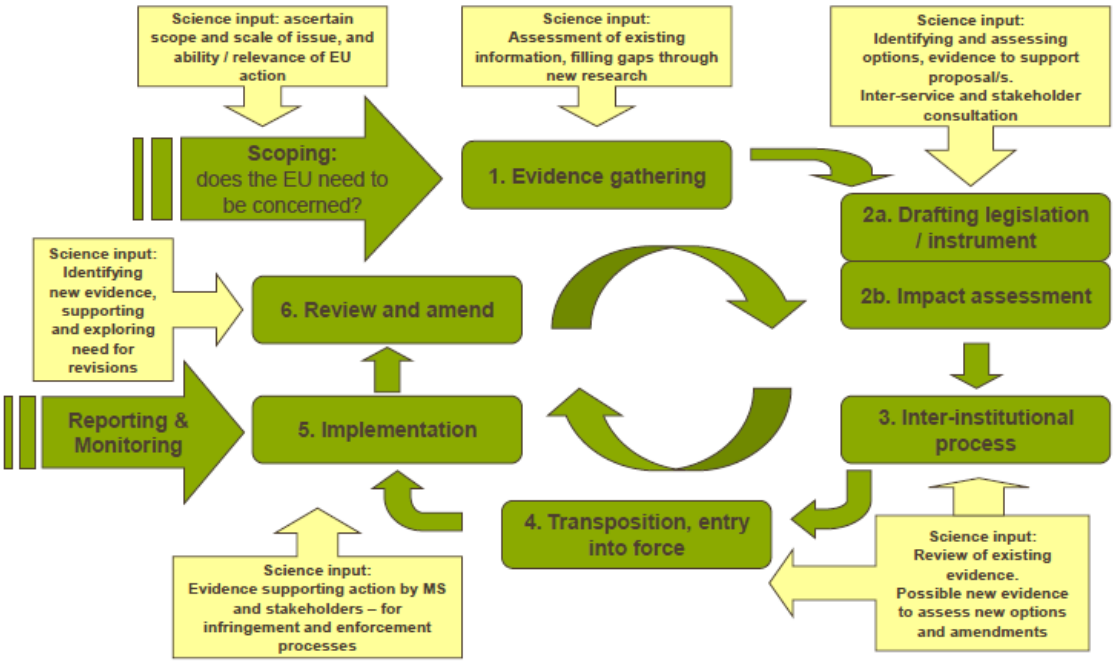


Figure 4.14: European Commission policy cycle (adapted from Zamparutti et al, 2012)

Table 4.1: Summary of key themes emerging from the evaluations - similarities and differences among embedded case studies

Category of intervention (cases)	Clarity of intervention objectives	Clarity of evaluation objectives	Context in which evaluation tool place	Project management context	Methods/ approaches used by evaluation	Evaluation framework	Evaluation use
National policy interventions	Some *ambiguity* around outcomes	**Flexible** – often evaluation the first time an explicit theory of change is developed for the intervention; hence little understanding of unintended consequences Lack of explicit options appraisal	Time, money and politics – drivers for policy interventions can be political and budgetary. Short timescales. **Evolving** policy	Internal staff changes common – lack of institutional memory	Mixed methods Qualitative with qualitative indicators.	Logic models;	Other than formative evaluation little evidence of what happened to them
EU policy interventions	**Clear objectives** – long period of negotiation before agreed legislation, results in a clear policy framework and theory of change.	**Rigid** evaluation framework – Better Regulation etc. sets core themes/questions	**Stable** and strong policy – policy context out with domestic policy/budgetary/elect oral Cycles. Long timescales. Different perspectives of multiple interests/ stakeholders	Internal staff changes common Top down hierarchy Lack of consensus on evaluation objectives etc. if multiple perspectives not well handled by commissioning project manager.	Mixed methods Long timescales allow for standardised monitoring/evaluation and reporting (e.g. ex ante, mid-term, ex post evaluations). Data allow for quantitative/modelling methods Quantitative and qualitative surveys of MSs common	Key research questions and Better Regulation informed evaluation framework	By their nature EU policy evaluations tend to be tailored to Instrumental use, and conceptual use.
Programme level (policy) interventions	**Lack of clarity** a common feature Programme/project relationship – lots of variability; sometimes deliberate discretion in implementation > inconsistent monitoring, optional indicators, own questions.	**Overambitious evaluation objectives and lack of explicit theory of change;** hence little understanding of unintended consequences. ToC makes impacts more explicit and easier to set evaluation objectives. Lack of options appraisal	**Evolving policy**. Work needed to make governance work, e.g. managing stakeholders' perspectives and input where multiple interests. Because policy is evolving potentially there is more chance for evaluation to influence, or else the policy is terminated or	Internal staff changes common – lack of institutional memory If not consensus in steering group – problematic. Proactive project manager helps. Larger number of stakeholders needs more management. Time (for relationship	Mixed methods Qualitative - interviews, Quantitative – surveys Existing data where available. **Programme/project level challenges**	Evaluation framework developed by evaluation team. Often more process and outcome focused, given long timescales to impacts (e.g. in relation to biodiversity)	Limited evidence of where influential, but usually for strategic or conceptual use, e.g. follow-up policy (as in NIAs) or biodiversity offsetting (which contributed in part to no further development of the policy)

			changed before evaluation reports. Policy can be highly mobile (e.g. biodiversity offsetting) – political drivers. But interventions often pilots or short lived.	building) – lack of time may say something about evaluation and impact/learning (and perceived value of evaluation by commissioners).			
Programme level (non-policy) initiatives	**Lack of clarity** a common feature Doing 'things' rather than implementing policy – in a policy vacuum and so unclear how evaluation feeds back into policy.	**Lack of explicit theory of change;** hence little understanding of unintended consequences. ToC makes impacts more explicit and easier to set evaluation objectives. Lack of options appraisal	**Interventions often pilots or short lived;** even if continue (e.g. partnership arrangements) may be expedient (e.g. means of securing funding) rather than policy driven.	Internal staff changes common – lack of institutional memory. As programme level policy interventions	Mixed methods Qualitative - interviews, Quantitative – surveys Existing data where available. **Programme/project level challenges**	As programme level policy interventions.	Limited knowledge of use of evaluation; accepted by client and contributed to conceptual/strategic/ process use

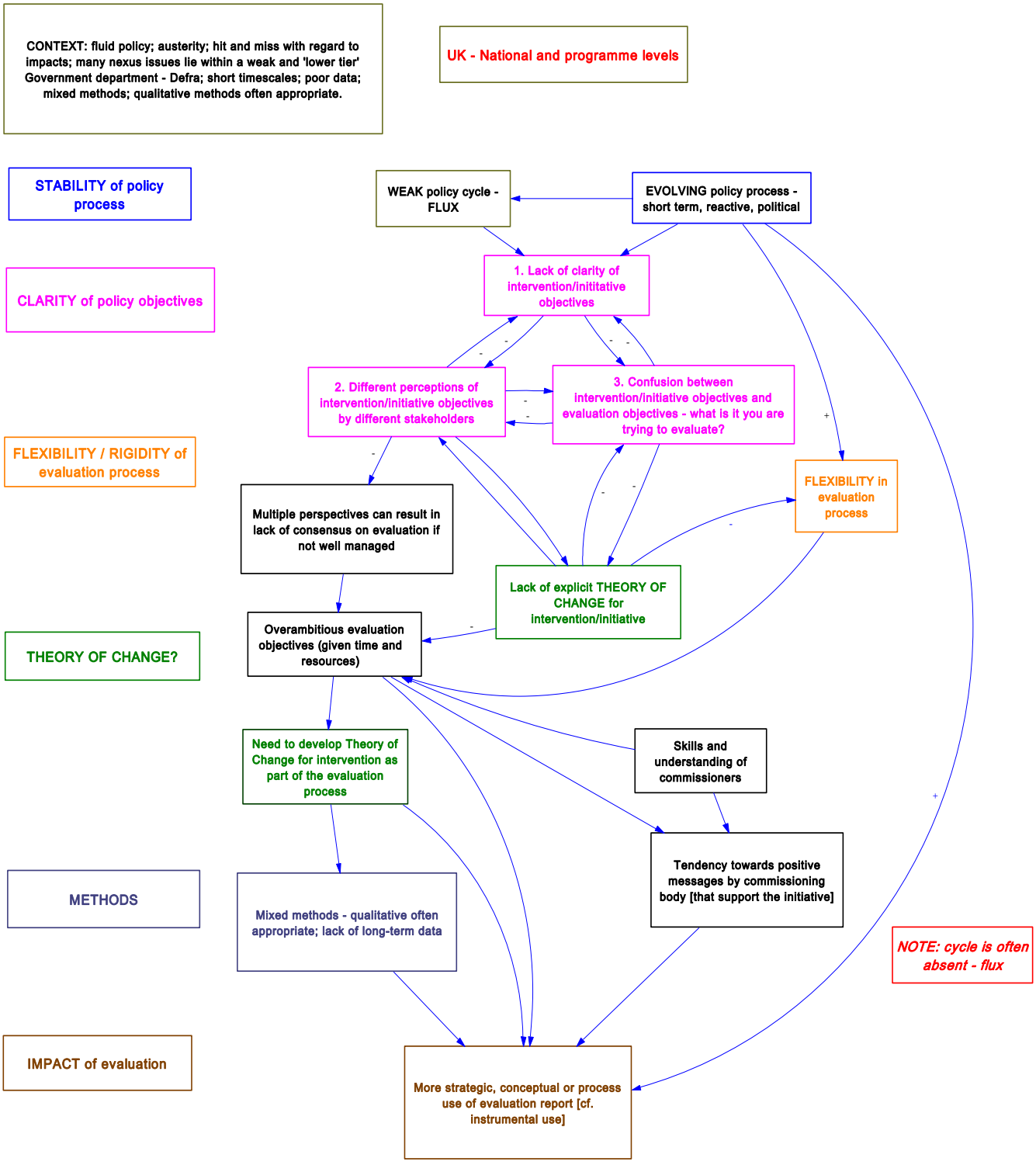


Figure 4.15: Complex mapping of emerging themes from evaluations of UK/National and programme level policy interventions and initiatives

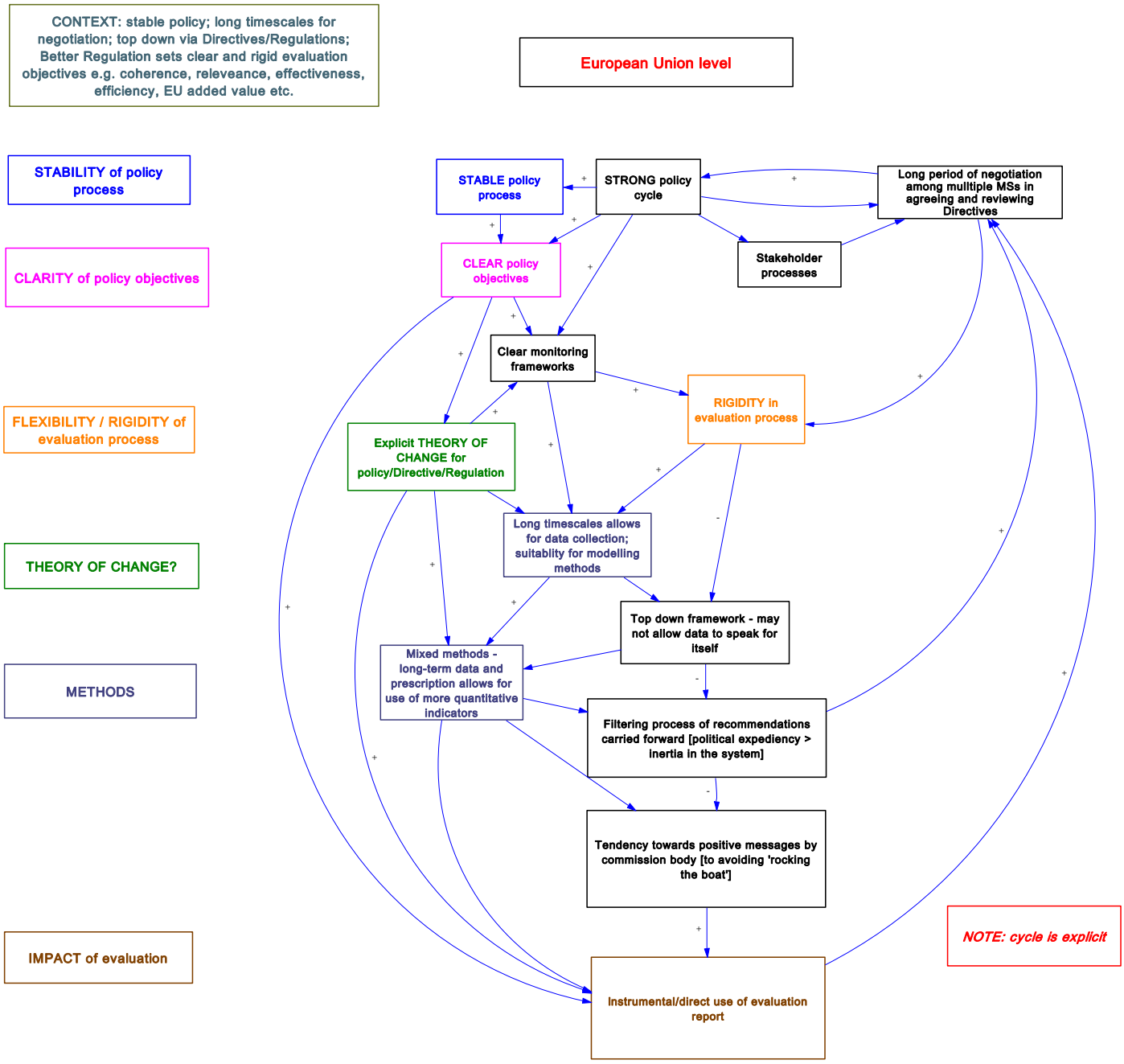


Figure 4.16: Complex mapping of emerging themes of Evaluations of EU policy interventions

5. Discussion: answering the meta-evaluation questions

There were four meta-evaluation questions

- Were the evaluations fit for purpose, and was their purpose clear? What lessons can we learn about assessing the effectiveness of the policy interventions?
- How has the framing of the evaluation been more or less useful for understanding complexity (e.g. logic model, objectives led)?
- Types of methods for types of complexity? What methods have been used for dealing with aspects of complexity found within environmental policy? For example, long term nature of impacts, interrelationship of social and physical systems? Which methods appear to have been most effective? Were some methods and techniques more suited to certain types of complexity?
- What factors lead to an evaluation being more (or less) influential of policy changes / outcomes / evaluation use?

This section presents the analysis carried out in order to answer these questions drawing out key issues and illustrating with some case examples (in boxes).

5.1 Were the evaluations fit for purpose, and was their purpose clear? What lessons can we learn about assessing the effectiveness of the policy interventions?

The main distinction found in terms of fitness for purpose was between UK and EU evaluations, and in particular UK programme level initiatives - because of the absence of a clear policy context - among those CEP has evaluated appear less likely to have an explicit theory of change already articulated. Some of these more exploratory interventions have learning and process as defining characteristics and require more attention to dialogue among stakeholders to avoid inconsistent evaluation objectives. Even policy interventions may lack an explicit theory of change and/or unclear objectives e.g. NIAs – even though intended to implement NEWP ecological networks they were given multiple and varying objectives that enabled a diversity of partnerships to be established, but added to the complexity in evaluation.

In general, the evaluations were fit for purpose inasmuch as they ended up often being tailor made because of the evolving policy context and the need for flexibility in establishing and modifying evaluation objectives. But evaluation cannot substitute for a strong policy process or clear policy purpose; it can question the policy/policy intervention, but it is only one element among many that determines policy.

In order to examine fitness for purpose in detail we considered questions around the clarity of both the policy and evaluation objectives, the circumstances in which the evaluation took places in terms of governance and stability of policy context together with the project management context. The focus on these aspects was expressed in our four working hypotheses:

1. Clarity and consensus on the objectives of the intervention support effective evaluation (e.g. due to consistent implementation in different projects under a programme).
2. Clarity and appropriateness of as well as consensus on the objectives of an evaluation will support efficacy.

3. A stable intervention governance and / or policy context will facilitate more effective evaluation.
4. Effective and efficient project management and governance leads to smoother and more effective evaluation.

Clarity and consensus of objectives of the policy/intervention and of the evaluation

In relation to these two aspects a number of issues emerged between the different case categories. With respect to NP policy/intervention objectives were clear but often lacked specific outcomes/targets, for example: the objective might be to "provide a positive experience" but there would be a lack of specific targets e.g. 20% more people stating that they feel better because of the intervention. Linked to this, within the PPI cases it was found that there was sometimes a lack of clarity on "how" the objectives were expected to be achieved, with different interpretations at the individual project level. EU policies on the other hand reflect objectives that are agreed by all parties i.e. Member States, and may represent compromises and include a degree of flexibility / ambiguity relating to national implementation, but this is within a stable policy context. This can be challenging for evaluation of EU level policy interventions, as the diversity of national implementation adds additional complexity to what are often already complex interventions.
At the level of the PI cases some interventions had vaguely defined objectives or in some cases the objectives weren't discussed until the launch of the evaluation. In projects where the objectives of the policy interventions were clearly defined this led to clear evaluation objectives.

Key Message: Interventions of an exploratory nature where learning and process are the defining characteristics represent ambiguity to the pathway of implementation. This requires more attention from stakeholders to avoid vaguely defined/inconsistent objectives.

Key message: It is important to ensure objectives of a policy/intervention are linked to a clear baseline and that there are specific measurable outcomes that an evaluation can then assess.

Evaluation of BBSRCs bioenergy public dialogue project, where:

- there was some lack of clarity among BBSRC about what the dialogue was actually trying to achieve

- and (perhaps related to the above point) the evaluation identified that in some events held as part of the dialogue as many as 15% of participants were not clear about the purpose (objectives) of the event/s.

NIAs: The objectives were very loosely framed and interpreted differently across the 12 participating NIAs. While this flexibility was considered a success factor it made clear comparison between the NIAs a challenge especially when coupled with a lack of clear targets for the outcomes.

In relation to the clarity of evaluation objectives, for the NPs generally they were set by a contracting authority although there was sometimes scope to clarify and refine the evaluation objectives. For the EU projects the evaluation (or assessment) objectives were set by the Commission, typically covered a limited number of specific research questions and were all considered clear, but being set at the outset sometimes had limited scope for review and revision. EU policy intervention evaluations generally followed formal review procedures which can limit the ability of an evaluation to set

'ambitious' objectives and/or develop new / amend proposed evaluation objectives. This may simplify the process, but a lack of flexibility may mean that aspects of complexity are overlooked.

From analysis of the PPI projects some had clearer evaluation objectives which were often set by the contractor and had the scope for revision at the beginning of the project which was particularly helpful when the evaluation objectives were overly ambitious which was observed on a number of occasions. Perhaps unsurprisingly, given what was observed in relation to the clarity of PI objectives, typically vaguely defined intervention objectives were shown to lead to uncertainties on the evaluation objectives and disagreements on how these objectives should be achieved. As with the PPIs, overambitious expectations of the evaluation were noted and it was observed that this was linked to initiatives with multiple stakeholders who had lack of consensus on priorities and implementation. However, it was seen that undefined intervention objectives could still produce clear evaluation objectives as in the case of evaluating a "process" rather than outcomes.

In general, having scope to refine evaluation objectives was considered useful especially for PPIs and PIs. Interestingly, due to their nature Strategic Environmental Assessments (SEA) tend to assess against wider environmental objectives and not just whether the plan meets its own objectives. The built in scoping phase of an SEA means objectives are automatically refined and agreed upon.

National Policy

The **Climate Change Strategy for Wales** and **BBSRC bioenergy public dialogue** had evaluation objectives set at the outset by the contracting authority, and these were felt to be clear and appropriate. The **Low cost resilience to flooding** had evaluation objectives developed by the evaluation team as part of the project. By discussing and revising the low cost resilience objectives over the initial phase of project implementation, it was possible to ensure these objectives were aligned with an appropriate aspect of the project and this ongoing interaction was important (because the project was an action research project) as objectives did change over time and the evaluation needed to reflect that.

Programme level Policy Interventions

FAW review: Initial objectives (set by the client) were adjusted based on the availability of data and time
NIAs: Objectives were clear and appropriate but a little ambiguous because they didn't set out how you might measure progress or indicate what success would look like. We struggled with what we were evaluating against.
Ex-post flood: a couple of the initial objectives were dropped due to lack of data established after the project started - need to allow time to have these discussions

Key message: Scope to discuss, amend and agree evaluation objectives as part of the initial work on an evaluation helps ensure clarity and fitness for purpose, and ongoing reflection on evaluation objectives is important especially when the policy objectives may be evolving over the time of the project

Key message: Setting clear programme level objectives at the outset to reflect the relationship between the programme and project level can aid robust evaluation

Key message: Full impact evaluation may not be possible for some complex policy interventions, especially where these are delivered over relatively short timescales. Scoping during the policy design phase what is possible for an evaluation to deliver would be helpful.

Stability of governance/policy context and the role of effective project management

With respect to stability of governance/policy context the EU context stood out as being most stable with clear policy cycles into which the evaluations fitted as is shown by the complexity mapping. What this leads to is an ability to "get the evaluation done", but perhaps less room for innovation and challenge to the existing policy because of the scope of the evaluation being quite prescribed and narrow. The particular context of EU policy interventions (i.e. Member State level involvement and implementation) means that evaluations may be limited in their ability to explore and propose more 'radical' outcomes and recommendations. From our analysis of the EU evaluations the issue of multiple, often numerous and sometimes conflicting stakeholders was less evident but tensions may exist between EU level governance of the policy intervention, and national (MS) governance in implementation.

EU level

The study on the application and effectiveness of the SEA Directive (Directive 2001/42/EC), was undertaken as part of the periodic (7 yearly) assessment of implementation of the SEA Directive, as required by the Directive itself.

The situation across the other three case categories was less clear with some examples of where there was a stable policy context e.g. ex-ante evaluations and SEAs that are informed by relevant EC and national level legislation as well as statutory and non-statutory SEA guidance. They offer a stable policy cycle ex-ante/mid-term/ex-post. Beyond this the policy contexts examined were often in development and/or in flux, see the example of the NIAs as a PPI example below.

PPI

A change in policy can mean a change of emphasis for the evaluation as was the case for NIAs where during the 'pilot' it became apparent that there wasn't going to be any further funding, which meant that the original objective of 'informing future integrated land-use and management initiatives' became a bit detached and there was a move towards showing contribution to the national growth.

With respect to the PPIs the governance arrangements were very much dependent on the extent of the novelty of the policy, pre-existing arrangements/links/ partnerships to build on and the approach of the evaluation itself. An issue that arose specifically within the PPIs and PIs was that of project boards or steering groups with multiple stakeholders and sometimes a lack of consensus on the objectives of the evaluation or the project. If this was coupled with project management issues (or changes in project management staffing) on behalf of the client then it could lead to considerable time spent gaining consensus on objectives or the evaluation being pulled in different directions.

The size of steering group appeared to be linked to the scale of the project but also the policy itself and whether it involved multiple actors. For the PIs numbers in a steering group ranged between one and twelve members, with five being an average. While a small group may indicate agreement is reached more easily, a very small group risks missing some perspectives that might have been useful.

Our analysis suggests that the effectiveness of the project manager (PM) and the working relationship can significantly affect the evaluation work; for example, a good PM can offer guidance, clarifying the project direction and facilitate the interaction between the SG and the evaluation team (e.g. collating SG comments and offering direction when SG members express different opinions). Without this, the evaluation work can be delayed and not meet its objectives so clearly.

PPI

NIAs: The steering group at times provided contradictory messages about priorities and direction and provided a considerable number of comments with little to no agreement/consensus. SG included people from different organisations with clear differences in perspective and some tensions (politically rooted) that manifested themselves through the project taking up time and resource for the evaluation.

LUS had a steering group of 11 people, who had fairly similar perspectives, while the PM collated comments from SG, thus facilitating the evaluation.

Key message: Complex policy interventions often require the involvement of diverse stakeholder groups, which means that different expectations, roles and views on objectives and progress will need to be considered and time needs to be allocated to getting agreement on objectives and evaluation.

Key messages: Time is required to develop a good working relationship with the PM to ensure that any issues around contrasting views on project boards are managed. Time available may be affected by tight project timeframes.

4.2 How has the framing of the evaluation been more or less useful for understanding complexity?

What was clear from the analysis was that no one framework was used exclusively across the CEP evaluation projects. Each project fitted into between 2 and 5 types of evaluation categories. This stems in part from the origins of CEP's evaluation work, which comes out of having expertise in nexus topics rather than being solely evaluation experts, and in part because of the type of evaluation requested by the clients.

Overall, the use of logic models has been widespread in the CEP sample and generally more explicit in recent years with the emphasis on the Magenta book being specified in tenders. Policies, however, are often lacking an explicit theory of change and the evaluation may be the first time such a theory of change has been articulated. Long term impacts, e.g. in relation to biodiversity, flooding are not capable of being evaluated within typical timescales for evaluations (2-3 years). Therefore, an emphasis on outcomes as the focus becomes necessary and a theory of change to understand how outcomes relate to long term intended impacts.

With respect to specific frameworks their usefulness (and related tools) in addressing specific aspects of complexity are presented in Table 5.1

Table 5.1: Frameworks and complexity

Evaluation Frameworks	Aspects of complexity	Examples
Theory based evaluation: use of logic models	Lack of clarity in the causality between actions and impacts (difficulty in attributing causality Degree of flexibility or tailoring / changes in the policy during implementation Timescales over which impacts might occur Availability of information and monitoring data relating to impacts	Flood Resilience Community Pathfinder Monitoring and evaluation of Phase 2 of the Nature Improvement Areas (NIA)
Formative evaluation	Level of unpredictability in the problem (e.g. 'Tipping points' - Sudden, unexpected changes due to a small change having a big impact & Technological advancements – improvements in abatement technology) Degree of flexibility or tailoring / changes in the policy during implementation Timescales over which impacts might occur Availability of information and monitoring data relating to impacts	Evaluation of Supporting the Uptake of Low Cost Resilience for Properties at Risk of Flooding
Participative evaluation	Multiple agencies / actors / stakeholders involved or targeted by the policy (may include conflicting interests)	Evaluation of Biodiversity Offsetting Pilot Phase
Quasi-experimental (matching or reflexive comparison)	Lack of clarity in the causality between actions and impacts (difficulty in attributing causality	Ex-Ante Evaluation and Strategic Environmental Assessment (SEA) of the Wales Rural Development Plan (2007-2013)

Key Message: In designing an evaluation it is important to recognise that timescales of delivery (activities and outputs) may differ from intervention outcomes and impacts, and that many impacts, especially in natural environment initiatives, cannot be detected over time periods of less than 5 years and in some cases decades. Where possible, therefore, longer-term monitoring should build on existing data and plan for the re-assessment of key indicators after the funded intervention has completed.

Key message: An effective evaluation is likely to require an evaluation framework supported by, for example, a clear logic model. Given the potential for delays between activities and outcomes and impacts a theory of change model(s) is a useful approach, accompanied by mechanisms for testing/validating the theory of change.

4.3 Types of methods for types of complexity?

What methods have been used for dealing with aspects of complexity found within environmental policy? Which methods appear to have been most effective? Were some methods and techniques more suited to certain types of complexity? Table 5.2 shows the range of methods used across the different evaluation projects carried out by CEP. What is clear from that is:

1. All projects use a mixed-methods approach, that is, different data sources e.g. documents, interviews, surveys, etc. and frequently both quantitative and qualitative data.

2. Qualitative data (collected through interviews, expert advice, workshops) is more frequently used than quantitative data. Qualitative data focuses on description, explanation and in understanding the context in which impacts might be realised. A mixed method approach allows triangulation of data and helps capture the perspectives of different stakeholders in different depths as necessary. Further, a mixed approach can allow consistent monitoring and evaluation for some objectives and more flexible reporting to reflect local objectives.

The top five types of complexity identified across the projects were:

- *Problem-related complexity*:
  - Problem has multiple elements
- *Policy/Response-related complexity:*
  - Multiple agencies / actors / stakeholders involved or targeted by the policy (may include conflicting interests)
  - High degree of flexibility or tailoring / changes in the policy during implementation
- *Impact-related complexity:*
  - Lack of clarity in the causality between actions and impacts (difficulty in attributing causality)
  - Poor availability of information and monitoring data relating to impacts

Table 5.2 shows how a mixed-method and predominantly qualitative approach is suited to addressing these areas of complexity.

Table 5.2: Methods suitable for dealing with types of complexity

Area of complexity	Project example	Aspect of mixed methods that was helpful	Aspect of qualitative data that was helpful	Other strategies to help address the complexity issue
Problem has multiple elements	Scottish Government Strategic Environmental Assessment (SEA) Pathfinder Research Project	The observational nature of the research ensured that the research was a 'true' picture of each case study, while the case study approach offered in-depth understanding of the key barriers to good practice and the benefits arising from SEA implementation.	In-depth understanding of issues	Use of causal chains to unpack pathways to impact
Multiple agencies / actors / stakeholders involved or targeted by the policy (may include conflicting interests	Catchment Base Approach (CaBA): Monitoring and evaluation (Phase 2) and wider adoption of CaBA for the period 2013-15	Use of online surveys followed by comparative analysis to- compare new partnerships and the more mature partnerships; compare the answers by the type of respondent, in particular their organisation, role in the catchment partnership and role in other regional/local forums. Follow up interviews were used for clarification of responses.	Multiple agencies / actors / stakeholders involved or targeted by the policy (may include conflicting interests	Engaging stakeholders and/or those involved or targeted by the policy
High degree of flexibility or tailoring / changes in the policy during implementation	Evaluation of the Biodiversity Offsetting Pilot Phase	Three primary research methods were used: literature review; document analysis; and semi-structured interviews and follow on discussions. This variety ensured that there could be some comparison across the pilots although they were all very different.	Quantitative indicators would not be appropriate due to the small sample size and the non-random design of the pilot selection. A qualitative approach was therefore adopted	Self-reported data and locally specific indicators can play a useful role, however, such approaches require support and facilitation, and therefore resources, and may result in inconsistent data. The use of existing national datasets and centralised analysis where possible can support effective, robust and efficient evaluation at both programme and local levels.
Lack of clarity in the causality between actions and impacts (difficulty in attributing causality	Flood Resilience Community Pathfinder Evaluation	Qualitative data to help explain quantitative findings	Providing a deeper understanding of outcomes and their potential link to impacts	Use of causal chains to unpack pathways to impact
Poor availability of information and monitoring data relating to impacts	Enhancing ex-post evaluation of flood and coastal erosion risk management plans and schemes	Use of interviews to fill in gaps where there was no quantitative data	Benefits were able to be fully described and their importance expressed	

Looking across the four case categories, some observations about the relationship between methods and types of complexity can be made. Across the EU projects, the most commonly seen types of complexity were: multiple agencies / stakeholders involved; flexible implementation (e.g. between EU and MS levels); and availability of data/indicators. The most common methods used were: interviews; and steering groups / expert advice. This suggests that stakeholder led methods have been used to help address complexity in implementation, stakeholder numbers / diversity and where there are limited data/indicators.

For the NP projects there was limited evidence of an association between methods used and types of complexity. However, all three evaluations reviewed made use of interviews and were also characterised by complexity related to the availability of evidence / data related to impacts. The use of interviews (and surveys etc.) is a method that enables perceptions of change / impact to be gathered and assessed in the absence of data / indicators.

More so than in any other category of evaluations, the evaluation of PPI projects involved undertaking a literature review, and using steering groups or groups of experts to collect evidence. Surveys and observational data were rarely used, while workshops were more common than usual along with interviews. The latter as a choice of evidence collection methods, links to the identification of 'Multiple agencies / actors / stakeholders involved or targeted by the policy', as the most commonly identified complexity criterion in the PPI case category.

Finally, for the PI projects the available information doesn't provide a clear link between the methods used and the complexities indicated across the projects. However the most common types of complexity are characterised by "multiplicity of factors": e.g. multiple agents/actors; problem has multiple elements; multiple components included in the initiatives. Another commonly identified complexity was the lack of clarity in the causality between actions and impacts. The most common methods used were: interviews and surveys. This could indicate that in order to deal with the variety of actors/elements etc. approaching stakeholders was considered to be the best way forward.

Key message: Qualitative and mixed methods are well-suited to addressing complexity in nexus-related evaluations.

Key message: The use of existing national datasets and centralised analysis where possible can help support effective, robust and efficient evaluation at both programme and local levels.

Key message: Self-reported data and locally specific indicators can play a useful role, however, such approaches require support and facilitation, and therefore resources, and may result in inconsistent data.

Key message: Careful consideration is needed in the commissioning and design of bespoke IT systems for short-term policy interventions to ensure that they are proportionate and provide value for money, taking into account the design, maintenance implementation and support costs.

Key message: Explicit options appraisal in policy development (ex-ante assessment can help inform counterfactual analysis (ex post) providing clear linkage between the different types of assessment/evaluation.

4.4 What factors lead to an evaluation being more (or less) influential of policy changes / outcomes / evaluation use?

The complex mapping shows the congruence of a number of factors, but the existence of a strong or weak policy cycle and stable/evolving policy appears critical, i.e. the evaluation needs to have somewhere to go – to feed into (as in EU policy cycle) if it is to have an influence – or at least

instrumental use. Otherwise the extent to which the evaluation has any influence is dependent on more arbitrary factors, e.g. the interest of a minister in a particular policy, change of policy priorities etc. and subject to the vagaries of an evolving policy in flux, under a system stewardship-type model.

Figure 4.11 above highlighted the low level of instrumental use of evaluation in UK programme initiatives and policy interventions and the high proportion of strategic use. That does not mean that evaluations are not being used, just that strategic use – for accountability and defending/promoting policy – may imply that evaluations were used, where they provide the appropriate answers ,to support policy development, or where they do not may be used as part of the rationale for dropping a certain policy direction or intervention, though it may actually have been for a range of other political or budgetary expediency or purposes, e.g. NIAs, biodiversity offsetting. Both of these hit the buffers as policy interventions – NIAs because of lack of funding and biodiversity offsetting because it became a political hot potato/non-starter. In both cases the evaluations were equivocal – at best they were lukewarm, identifying only marginal benefits and in the case of biodiversity offsetting considerable costs and risks. In both cases policy was highly fluid – examples of 'system stewardship' perhaps (cf. a policy cycle). Other examples of evaluation use are shown in Table 5.3.

Table 5.3: Examples of evaluation use

Example	Types of evaluation use
SEA of the Wales RDP: The proposed mitigation measures were to be considered as part of the eligibility criteria for rural development schemes and the conditions put in place for receiving payment. The WG fully supported that SEA monitoring would best be integrated into the monitoring of the implementation of the RDP and the development of the monitoring framework was to ensure that this was taken forward.	Strategic use Process related use
The NIA programme came to an end with the end of the funding in 2015. The lessons learned from the NIA evaluation though were used in the scoping of the Countryside Stewardship facilitation fund for Defra. So the NIA evaluation informed the design of a new policy. Groups formed from four of the NIAs were awarded funding in the first round of the CSFF. Some of the NIAs also reported that they had used the evaluation outcomes in proposals and had succeeded in securing funding from other sources. The NIA evaluation found that the Online reporting tool (used by NIAs to record their monitoring data) posed challenges and some users struggled to operate the tool independently even though guidance, training and support were provided. On the basis if this Natural England decommissioned the tool.	Conceptual use Strategic use Process related use
The 2007 EU Directive on pyrotechnic articles was repealed and replaced with a modified Directive in 2013. As the evaluation was completed in 2011, it is assumed that the evaluation (and wider assessment it was part of) formed part of the evidence base used by the Commission to inform revisions to the 2007 Directive.	Instrumental use
The study on the **application and effectiveness of the SEA Directive (Directive 2001/42/EC**) was intended as the main input to the Commission's second report on the application and effectiveness of the SEA Directive, due in 2016. This work is ongoing, and the formal Commission effectiveness report will be published in due course, after a process of consultation through MS expert working groups / more widely. The report and recommendations may also be used to guide the Commission in the preparation of issue/topic specific guidelines for MS, e.g. around 'tiering' between EIA and SEA etc.	Instrumental use Process related use
Evaluation of the catchment-based approach - pilot stage The evaluation team worked closely with Defra and the Environment Agency to ensure that lessons learnt throughout the pilot phase informed the development of the policy framework for the wider adoption of the Catchment Based Approach. The team provided endorsement for the scheme to continue on a national scale which was taken on board as CABA Phase 2 was launched. After this evaluation a catchment based approach was established and funding is continued. Rivers Trust (one of the stakeholders of the pilot stage project) have since taken an established secretarial role in this initiative.	Instrumental use Strategic use Conceptual use Process related use

Key message: High level of instrumental use is seen in EU policy evaluations, because they are designed to deliver that within a strong policy cycle.

Key message: Much of UK environmental policy making exhibits a high degree of flux – more typical of a system stewardship model of policy making/governance than a typical policy cycle. Consequently, evaluation has to be more nimble and flexible to respond to ongoing changes in policy purpose, design and implementation.

Key message: Evaluation can have influence in a more indirect way – conceptual, strategic or process influence and these are more likely in a system stewardship model of policy making.

Key message: An important human factor that influenced an assessment of evaluation use was minimal post-evaluation interaction with evaluators, due to the contractual nature of the projects reviewed in this study.

6. Conclusions and next steps

6.1 Overview of CEP meta-evaluation

A summary of the findings of the meta-evaluation is provided here from the key messages in Chapter 5.

1. Fitness of purpose of evaluations

- In general the evaluations were fit for purpose, but often because they were tailor made to the circumstances.

Clarity and consensus on policy objectives

- Interventions of an exploratory nature where learning and process are the defining characteristics represent ambiguity to the pathway of implementation. This requires more attention from stakeholders to avoid vaguely defined/inconsistent objectives.
- It is important to ensure objectives of a policy/intervention are linked to a clear baseline and that there are specific measurable outcomes that an evaluation can then assess.

Clarity, consensus and appropriateness of evaluation objectives

- Scope to discuss, amend and agree evaluation objectives as part of the initial work on an evaluation helps ensure clarity and fitness for purpose, and ongoing reflection on evaluation objectives is important especially when the policy objectives may be evolving over the time of the project
- Setting clear programme level objectives at the outset to reflect the relationship between the programme and project level can aid robust evaluation
- Full impact evaluation may not be possible for some complex policy interventions, especially where these are delivered over relatively short timescales. Scoping during the policy design phase what is possible for an evaluation to deliver would be helpful.

Stability and effectiveness of policy/governance context

- Complex policy interventions often require the involvement of diverse stakeholder groups, which means that different expectations, roles and views on objectives and progress will need to be considered and time needs to be allocated to getting agreement on objectives and evaluation.
- Time is required to develop a good working relationship with the PM to ensure that any issues around contrasting views on project boards are managed. Time available may be affected by tight project timeframes.

2. Framing of the evaluation and complexity

- In designing an evaluation it is important to recognise that timescales of delivery (activities and outputs) may differ from intervention outcomes and impacts, and that many impacts, especially in natural environment initiatives, cannot be detected over time periods of less than 5 years and in some cases decades. Where possible, therefore, longer-term monitoring should

build on existing data and plan for the re-assessment of key indicators after the funded intervention has completed.

- An effective evaluation is likely to require an evaluation framework supported by, for example, a clear logic model. Given the potential for delays between activities and outcomes and impacts a theory of change model(s) is a useful approach, accompanied by mechanisms for testing/validating the theory of change.

3. Methods and complexity

- Qualitative and mixed methods are well-suited to addressing complexity in nexus-related evaluations.
- The use of existing national datasets and centralised analysis where possible can help support effective, robust and efficient evaluation at both programme and local levels.
- Self-reported data and locally specific indicators can play a useful role, however, such approaches require support and facilitation, and therefore resources, and may result in inconsistent data.
- Careful consideration is needed in the commissioning and design of bespoke IT systems for short-term policy interventions to ensure that they are proportionate and provide value for money, taking into account the design, maintenance implementation and support costs.
- Explicit options appraisal in complex policy development (ex-ante assessment can help inform counterfactual analysis (ex post) providing clear linkage between the different types of assessment/evaluation.

4. Factors affecting influence of evaluation

- High level of instrumental use is seen in EU policy evaluations, because they are designed to deliver that within a strong policy cycle.
- Much of UK environmental policy making exhibits a high degree of flux – more typical of a system stewardship model of policy making/governance than a typical policy cycle. Consequently, evaluation has to be more nimble and flexible to respond to ongoing changes in policy purpose, design and implementation.
- Evaluation can have influence in a more indirect way than instrumental – i.e. through conceptual, strategic or process influence and these are more likely in a system stewardship model of policy making than instrumental.
- An important human factor that influenced an assessment of evaluation use was minimal post-evaluation interaction with evaluators, due to the contractual nature of the projects reviewed in this study, i.e. it is often unclear what happens to evaluations.

6.2 Strategies for dealing with complexity in evaluations

What we might call *'intrinsic'* complexity (because of the subject matter, i.e. issue and impact related) exists equally for nexus-related policy interventions at EU and national/regional/local levels. The key difference, from the meta-evaluation, appears to be the strength or dynamism of the governance and policy making context, which dictates the '*extrinsic'* complexity – that exerted not by the subject matter but by the complex web of interrelationships of stakeholders and processes (i.e. policy/response related complexity). In the UK examples, where policy making is more in flux – because of increasing devolved responsibilities to multiple stakeholders – this extrinsic complexity is enhanced. In the EU, it is systematised through rigid processes and frameworks. Consequently evaluations of policy interventions where policy is in flux have to deal with very different contexts and enhanced complexity compared to those where there is a strong policy cycle and an evaluation's purpose is not only clear, but explicitly prescribed.

While the evaluations examined were largely fit for purpose those that were less straightforward were the ones without a clear policy framework within which to work (i.e. programme level initiatives). This also makes it difficult to understanding how the evaluation was used (if at all) since the policy context is absent. From the EU to national policy interventions down to programme level interventions and initiatives there was a decreasing level of clarity regarding policy and evaluation objectives, reflecting the weak policy cycle (or system stewardship) context for those evaluations.

The extent to which framing of the evaluation affects how complexity is addressed in practice comes down to the fact that evaluators do what they are asked to do by commissioners of evaluations – there may be limited scope to bring in novel approaches or methods. The methods being used are therefore the ones clients are familiar with and understand and that can be used readily for quick evaluations on small budgets; in our case the ones we are experienced at using and include in tender proposals and that are accepted by clients. Since a large proportion of evaluations within our meta-evaluation were for relatively small budgets, qualitative methods and theory based approaches (theory of change, for example) were most appropriate, especially where the evaluation may have been the first time an explicit theory of change had been elaborated.

All of the above considerations have implications for commissioners of evaluations in relation to complex policy interventions around the nexus. Commissioners and evaluators need to be aware that an assumption of a traditional policy cycle (however fuzzy that may have been in practice) may no longer be appropriate and that evaluation therefore is less likely to have direct, instrumental use than might have been anticipated. Rather than a fuzzy policy cycle, if policy is in a constant state of flux (system stewardship) the purpose, design and implementation of policy are all potentially moving targets, which make it harder to pin down evaluation objectives than when the purpose (objective) of policy is clear.

This meta-evaluation provides substantive evidence of this type of policy flux and the challenges the evaluations in those situations faced along with the need to tailor-make evaluations each time to those circumstances. For such evaluations to have impact increasingly evaluators will need to be nimble and responsive to changing policy purpose, design and implementation and understand where within this new system stewardship 'policy triangle' evaluation could impact most effectively.

6.3 Key questions for new evaluations

1. What is the nature of the policy context in which your evaluation is being carried out? Would you describe it as evolving, stable, unclear, high profile?
2. How far are the objectives of the policy/intervention/initiative clear and amenable to evaluation? Are the expected outcomes and impacts clear?
3. How far are the objectives of the evaluation clear and achievable given the nature/timing of the policy/intervention/initiative and the resources of the evaluation?
4. Are there multiple stakeholders involved as part of the steering group for the policy intervention/initiative? How far is there consensus across perspectives? Are their clear mechanisms in place to enable management of different perspectives?
5. Is there a clear and active Project Manager for the evaluation?
6. What are the expectations of the client in relation to the ability of the evaluation to evaluate longer term impacts?
7. What types of complexity are most relevant to the evaluation? [refer to the four categories and sub-categories]
8. To what extent do you think your methods are appropriate for evaluating these complexities? What strategies can you use to address these specific aspects of complexity?
9. What types of impact are expected by your evaluation? How will the client assess if they have been realised?
10. How can you improve the impact of your evaluation? Where are the points of influence within the evaluation?

References

BetterEvaluation. (2013). Review Evaluation (do meta-evaluation) http://betterevaluation.org/en/plan/manage/review_evaluation

Campbell, R., Pound, P., Morgan, M., Daker-White, G., Britten, N., Pill, R. & Donovan, J. (2011). Evaluating meta-ethnography: systematic analysis and synthesis of qualitative research. *Health Technology Assessment*, 15(43), 1–164. http://doi.org/10.3310/hta15430

Caracelli, V. J., & Preskill, H. S. (2000). *The expanding scope of evaluation use*. Jossey-Bass.

Carpenter, J., Rücker, G., & Schwarzer, G. (2011). Assessing the Sensitivity of Meta-analysis to Selection Bias: A Multiple Imputation Approach. *Biometrics*, *67*(3), 1066-1072.

Centre for the Evaluation of Complexity Across the Nexus [CECAN]. (2016). About CECAN. http://www.cecan.ac.uk/about-us

Davies, R. (2013). *Planning Evaluability Assessments, A Synthesis of the Literature with Recommendations*, 58. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/248656/wp40-planning-eval-assessments.pdf

Defra (2011). *Defra's Evidence Investment Strategy: 2010-2013 and beyond. 2011 Update*. Retrieved from https://www.gov.uk/government/publications/defra-s-evidence-investment-strategy-2010-to-2013-and-beyond-2011-update

Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. (2005). Synthesising qualitative and quantitative evidence: a review of possible methods. *Journal of Health Services Research & Policy*, 10(1), 45–53. http://doi.org/doi:10.1258/1355819052801804

Downe, J., Martin, S., & Bovaird, T. (2012). Learning from complex policy evaluations. *Policy and Politics*, 40(4), 505–523. http://doi.org/10.1332/030557312X645766

European Commission [EC]. (2006). *Evaluation Methods for the European Union's External Assistance. Guidelines for Project and Programme Evaluation*, Volume 3. Reproduction (Vol. 3). Retrieved from http://ec.europa.eu/europeaid/evaluation/index.htm

European Commission. (2015a). *Guidelines on Impact Assessment.* Europa.eu. Retrieved from http://ec.europa.eu/smart-regulation/guidelines/ug_chap3_en.htm

European Commission. (2015b). *Regulatory Fitness and Performance Programme (REFIT) State of Play and Outlook "REFIT Scoreboard,"* (May), 1–149.

Fern, E. F., & Monroe, K. B. (1996). Effect-size estimates: Issues and problems in interpretation. *Journal of Consumer Research*, *23*(2), 89-105.

Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: effective use of theories of change and logic models* (Vol. 31). John Wiley & Sons.

Glouberman, S., & Zimmerman, B. (2002). Complicated and complex systems: what would successful reform of Medicare look like? *Romanow Papers*, *2*, 21-53.

Gough, D., Thomas, J. & Oliver, S. (2012). Clarifying differences between review designs and methods. *Systematic Reviews*, 1(1), 28. http://doi.org/10.1186/2046-4053-1-28

Hallsworth, M. (2011) *System Stewardship - The future of policy making?* London: Institute for Government. http://observgo.uquebec.ca/observgo/fichiers/85195_aeepp1.pdf

Hanssen, C. E., Lawrenz, F., & Dunet, D. O. (2008). *Concurrent Meta-Evaluation: A Critique*, 572–582.

Hargreaves, M. B., & Podems, D. (2012). Advancing Systems Thinking in Evaluation: A Review of Four Publications. *American Journal of Evaluation*, 33(3), 462–470. http://doi.org/10.1177/1098214011435409

HM Government. (2013). *What Works : evidence centres for social policy*, (March), i-9. Retrieved from www.cabinet-office.gov.uk

HM Treasury. (2003). *The Green Book : Appraisal and Evaluation in Central Government. Evaluation*, (October 2002), 118. http://doi.org/http://greenbook.treasury.gov.uk/index.htm

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

HM Treasury. (2012). *Quality in qualitative evaluation: a framework for assessing research evidence* (supplementary Magenta Book guidance)

Jaffe, A. B., Newell, R. G., & Stavins, R. N. (2005). A tale of two market failures: Technology and environmental policy. *Ecological economics*, *54*(2), 164-174.

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New directions for evaluation*, *2000*(88), 5-23.

Kromrey, J. D., Hogarty, K. Y., Ferron, J. M., Hines, C. V., & Hess, M. R. (2005). Robustness in meta-analysis: An empirical comparison of point and interval estimates of standardized mean differences and Cliff's delta. In *American Statistical Association 2005 Joint Statistical Meetings* (p. 7).

Levitt, R., Martin, S., Nutley, S., & Solesbury, W. (2010). *Evidence for accountability*. Retrieved from http://www.nuffieldfoundation.org/news/evidence-accountability-using-evidence-audit-inspection-and-scrutiny-uk-government

Mackay, K., Clark, M., Sartorius, R., & Bamberger, M. (2009). *Monitoring & Evaluation: Some Tools, Methods & Approaches*. World Bank, 26.

Magro, E., & Wilson, J. R. (2013). Complex innovation policy systems: Towards an evaluation mix. *Research Policy*, 42(9), 1647–1656. http://doi.org/10.1016/j.respol.2013.06.005

Pawson, R. (2013). *The science of evaluation: a realist manifesto*. Sage.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review--a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy*, 10 Suppl 1(July), 21–34. http://doi.org/10.1258/1355819054308530

Peck, L. R., & Gorzalski, L. M. (2009). An Evaluation Use Framework and Empirical Assessment. *Journal of Multi-disciplinary Evaluation*, 6(12), 139–156.

Pollitt, C. (2013). Ray Pawson (2013) The Science of Evaluation: A Realist Manifesto. *International Review of Administrative Sciences*, *79*(3), 580-582. http://doi.org/10.1177/0020852313497455

Preskill, H., Gopal, S., Mack, K., & Cook, J. (2014). *Evaluating complexity: Propositions for improving practice*. http://www.issuelab.org/resource/evaluating_complexity_propositions_for_improving_practice

Rogers, P. J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation*, 14(1), 29–48. http://doi.org/10.1177/1356389007084674

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

ScienceWise. (2013). *The Government's Approach to Public Dialogue on Science and Technology*, 8.

Sciencewise. (2016). *Quality in public dialogue: A framework for assessing the quality of public dialogue.* For the Department of Business, Innovation & Skills. (March), 1–26.

Scriven, M. (1969). *An Introduction to Meta-Evaluation*. Educ Prod Rep.

Snyder, S. (2013). The Simple, the Complicated, and the Complex: Educational Reform Through the Lens of Complexity Theory. *OECD Education Working Papers*, No. 96, OECD Publishing. http://dx.doi.org/10.1787/5k3txnpt1lnr-en

Stufflebeam, D., & Shinkfield, A. (2007). *Evaluation theory, models, & applications*. San Francisco: John Wiley.

The Nexus Network. (2016). *Nexus Shocks*. http://www.thenexusnetwork.org/projects/networking-grants/nexus-shocks/

The Oxford Handbook of Public Policy. (2014). *Journal of Policy Analysis and Management*, 33(4), 1036–1043. http://doi.org/10.1002/pam

United Nations Economic and Social Commission for Asia and the Pacific [UN ESCAP]. (2010) ESCAP M&E SYSTEM - *Monitoring & Evaluation System Overview and Evaluation Guidelines*. https://www.unescap.org/sites/default/files/ESCAP-M&E-System.pdf

Warburton, D., Wilson, R., & Rainbow, E. (2006). *Making a Difference: A guide to Evaluating Public Participation in Central Government*, 1–47. Retrieved from http://www.involve.org.uk/evaluation/

Weiss, C. H. (1981). Measuring the use of evaluation. *Utilizing evaluation: Concepts and*

measurement techniques, 17-33.

Westhorp, G. (2014). *Realist Impact Evaluation*. Research and Policy Development, (September).

What Works Network. (2014). *What Works? Evidence for decision makers*, (November).

Wong, G., Greenhalgh, T., Westhorp, G., & Pawson, R. (2014). Development of methodological guidance, publication standards and training materials for realist and meta-narrative reviews: the RAMESES (Realist And Meta-narrative Evidence Syntheses – Evolving Standards) project. *Health Services and Delivery Research*, 2(30), 1–252. http://doi.org/10.3310/hsdr02300

Zamparutti, T., White, O., Sheate, W., & Baker, J. *Assessing and Strengthening the Science and EU Environment Policy Interface*, Technical Report - 2012 – 059, European Communities, available at http://bookshop.europa.eu/en/assessing-and-strengthening-the-science-and-eu-environment-policy-interface-pbKH3112590/

Appendices

Appendix 1 – List of approaches in evaluation

Appreciative Inquiry

A participatory approach that focuses on existing strengths rather than deficiencies - evaluation users identify instances of good practice and ways of increasing their frequency.

Beneficiary Assessment

An approach that assesses the value of an intervention as perceived by the (intended) beneficiaries, thereby aiming to give voice to their priorities and concerns.

Case study

A research design that focuses on understanding a unit (person, site or project) in its context, which can use a combination of qualitative and quantitative data.

Collaborative Outcomes Reporting

An approach that builds on contribution analysis, adding expert review and community review of the assembled evidence and conclusions.

Contribution Analysis

An approach for assessing the evidence for claims that an intervention has contributed to observed outcomes and impacts

Critical System Heuristics

An approach used to surface, elaborate, and critically consider boundary judgments, that is, the ways in which people/groups decide what is relevant to the system of interest (any situation of concern).

Democratic Evaluation

Democratic Evaluation is an approach where the aim of the evaluation is to serve the whole community.

Developmental Evaluation

An approach appropriate for evaluations of adaptive and emergent interventions, such as social change initiatives or projects operating in complex and uncertain environments

Empowerment Evaluation

Empowerment Evaluation is an approach which provides communities with the tools and knowledge that allows them to monitor and evaluate their own performance.

Horizontal Evaluation

An approach that combines self-assessment by local participants and external review by peers

Innovation History

A way to jointly develop an agreed narrative of how an innovation was developed, including key contributors and processes, to inform future innovation efforts

Institutional Histories

An approach for creating a narrative that records key points about how institutional arrangements have evolved over time and have created and contributed to more effective ways to achieve project or or program goals

Most Significant Change

Collects and analyses personal accounts of change, includes processes for learning about what changes are most valued by individuals and groups.

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 Harvesting has proven to be especially useful in complex situations when it is not possible to define concretely most of what an intervention aims to achieve, or even, what specific actions will be taken over a multi-year period.

Outcome Mapping

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 via 'boundary partners'.

Participatory Evaluation

A range of approaches that engage stakeholders (especially intended beneficiaries) in conducting the evaluation and/or making decisions about the evaluation

Participatory Rural Appraisal

Enables farmers to analyse their own situation and develop a common perspective on natural resource management and agriculture at village level. (Recently renamed Participatory Learning for Action (PLA)

Positive Deviance

Involves intended evaluation users in identifying 'outliers' – those with exceptionally good outcomes - and understanding how they have achieved these.

Randomised Controlled Trials

An approach that produces an estimate of the mean net impact of an intervention by comparing results between a randomly assigned control group and experimental group or groups.

Realist Evaluation

Realist evaluation is a form of theory-driven evaluation but is set apart by its explicit philosophical underpinnings. Based in realist philosophy, it considers that interventions work (or not) because actors make particular decisions in response to what is provided by the intervention (or not).

Social Return on Investment

Identifies a broad range of social outcomes, not only the direct outcomes for the intended beneficiaries of an intervention.

Utilisation-Focused Evaluation

Uses the intended uses of the evaluation by its primary intended users to guide decisions about how an evaluation should be conducted.

Appendix 2

Table A2.1: Scale and Theme categories

Scale			
Local	**Regional**	**National**	**Multi-national**
Evaluation is focused on particular area/municipality/ community/ neighbourhood/district	Evaluation is focused on a recognised (subnational) geographical or administrative region.	Evaluation relates to a particular country as a whole. For the purposes of this categorisation national includes both UK wide and individual devolved administrations (Wales, Scotland, NI)	Evaluation is focused on a scale involving multiple countries (including EU level and multinational regions in EU or globally).
Theme			
Policy intervention	**Dialogue and engagement**	**Methodological**	**Programme/Initiative**
Evaluation focus on effectiveness of policies (including outcomes, necessity, efficiency, validity, etc.)	Evaluations focusing primarily on public or stakeholder engagement, including public dialogue.	Evaluation focus on methods and approaches, project planning and management arrangements, monitoring and data arrangements.	Evaluation is focused on programmes/ initiatives that can be but are not necessarily delivered under a policy

Table A2.2: Policy area

Natural environment	**Agriculture/ rural/forestry**	**Energy**	**Water management**	**Climate Change**	**Other / multiple**
Including policies and interventions related to biodiversity, ecosystem services, green infrastructure etc.			Including policies and interventions related to flood management and resilience as well as water quality and integrated water management.	Including mitigation and adaptation	Including policies or interventions in one or more sectors not included as specific (NEXUS related) categories, and cross-sectoral policies or interventions such as EU cohesion policy

Table A2.3: Types of evaluation

Formative	Summative	Developmental	Participatory	Theory based
Formative evaluation is generally any evaluation that takes place before or during a project's implementation with the aim of improving the project's design and performance.	Summative evaluation occurs at the end of a program cycle and provides an overall description of program effectiveness. It examines program outcomes to determine overall program effectiveness.	Developmental evaluation is an approach to understanding the activities of a program operating in dynamic, novel environments with complex interactions. It focuses on innovation and strategic learning rather than standard outcomes and is as much a way of thinking about programs-in-context and the feedback they produce.	Participatory evaluation is an approach that involves the stakeholders of a programme or policy in the evaluation process. This involvement can occur at any stage of the evaluation process, from the evaluation design to the data collection and analysis and the reporting of the study	Allows a much more in-depth understanding of the workings of a program or activity—the "program theory" or "program logic." Typically applies a logic model: Inputs - Activities – Outputs – Benefits – Outcomes
Ex-ante evaluation	**Ex-post evaluation**	**Experimental**	**Quasi-experimental (Matching or reflexive comparison)**	**Non-experimental**
An evaluation of a policy / intervention as part of its development or planning (i.e. before implementation).	An evaluation of the outcomes of the implementation of a policy / intervention (i.e. after implementation)	Provide a strong description of the counterfactual and hence effect of an intervention This design involves gathering a set of individuals (or other unit of analysis) equally eligible and willing to participate in the program and randomly dividing them into two groups: those who receive the intervention (treatment group) and those from whom the intervention is withheld (control group)	***Matching*** involves identifying non–program participants comparable in essential characteristics to participants. Both groups should be matched on the basis of either a few observed characteristics or a number of them that are known or believed to influence program outcomes. In a ***reflexive comparison***, the counterfactual is constructed on the basis of the situation of program participants before the program. Thus, program participants are compared to themselves before and after the intervention and function as both treatment and comparison group. This type of design is particularly useful in evaluations of full-coverage interventions such as nationwide policies and programs in which the entire population participates and there is no scope for a control group	Program participants can be compared to non-participants using statistical methods to account for differences between the two groups. Outcomes and impacts assessed without a conventional counterfactual

Table A2.4: Methods of data collection

Literature Review	Data/indicator review	Observation	Surveys / questionnaires	Developing case studies
A literature search enables the evaluator to make the best use of previous work in the field under investigation, and hence to learn from the experiences, findings and mistakes of those who have previously carried out similar or related work. A literature search can provide invaluable insight into the program area being evaluated and should, consequently, always be undertaken at an early phase of an evaluation study.	Data collection method aimed at discovering pre-existing data that can be used in the evaluation. A file review, seeks insight into the specific program being evaluated. Data already collected on and about the program and its results may reduce the need for new data, much as is the case in a literature search.	Direct observation generally provides more powerful evidence than that which can be obtained from secondary sources. Going into the "field" to observe the evaluation subject first-hand can be an effective way of gathering evidence. The results of field observation, recorded through photos or videos, can also be helpful and may have a powerful impact on the reader if used in the evaluation report.	Surveys are systematic ways of collecting primary data-quantitative, qualitative or both-on a program and its results from persons associated with the program	Case studies assess program results through in-depth, rather than broad, coverage of specific cases or projects.
Interviews	**Workshops / events**	**Steering group / Expert advice**	**Participant diaries**	**Interviews**
Face-to-face or telephone interviews allows spending time talking to people and provides a greater depth of evidence.	Workshops - a series of meetings emphasizing interaction and exchange of information among a usually small number of participants	Evidence acquired from steering group meetings or experts in the field providing comments /suggestions etc.	Tables, forms, online reporting tools used by participants to record their data, later to be used in quantitative and/or qualitative by the evaluator.	Face-to-face or telephone interviews allows spending time talking to people and provides a greater depth of evidence.

Table A2.5: Analysis approaches

Statistical analysis	Qualitative analysis	Analysis of case studies	Inductive analysis	Analysis of Further Program Results
Statistical analysis involves the manipulation of quantitative or qualitative (categorical) data to describe phenomena and to make inferences about relationships	A technique for systematically describing written, spoken or visual communication	Analysing a case typically includes the following steps: - Defining the issue(s)	The following are some of the purposes underlying the development of the general inductive approach. These	Analytically tracing the measured direct results to further impacts

among variables.		- Analysing the case data - Generating alternatives - Selecting decision criteria - Analysing and evaluating alternatives - Selecting the preferred alternative - Developing an action/implementation plan	purposes are similar to other qualitative analysis approaches. 1. To condense extensive and varied raw text data into a brief, summary format. 2. To establish clear links between the research objectives and the summary findings derived from the raw data and to ensure these links are both transparent (able to be demonstrated to others) and defensible (justifiable given the objectives of the research). 3. To develop of model or theory about the underlying structure of experiences or processes which are evident in the text (raw data). Inductive coding begins with close readings of text and consideration of the multiple meanings that are inherent in the text. The researcher then identifies text segments that contain meaning units, and creates a label for a new category into which the text segment is assigned. (Source: Thomas D.R. 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. American Journal of Evaluation 27: 237)	

Use of Models	Mapping / network analysis	Cost-benefit and Cost-effectiveness Analysis	Attribution analysis	
Simulation models: three main components: input data, a mathematical model and output data **Input-output model;** **Micro economic analysis** etc.	A method that can be used to help to describe ideas or data / other evidence in a visual or spatial form.	Cost benefit analysis involves the systematic identification and assessment of the benefits and costs of a particular intervention or policy. CBA typically considers monetary costs and benefits. Cost effectiveness analysis typically identifies and assesses the monetary costs of an intervention or policy and compares these to expected 'effects', which may not necessarily be monetised.	An assessment of how much of an observed change can be attributed to the intervention.	

Table A2.6: Types of complexity

Issue-related complexity	Policy/response-related complexity	Impact-related complexity
• problem has multiple elements, • variability in the physical characteristics of the area, • geographic scale of the problem, • sensitivity to socio-demographic characteristics of the area, • unpredictability in the problem	• multiple components included in the policy/programme/initiative, • multiple agencies/stakeholders involved or targeted by the policy, • high degree of flexibility or tailoring/changes in the policy during implementation	• multiple types/range of possible/expected outcomes and impacts, • unexpected/unintended impacts (positive/negative), • interactions between components of a policy, • lack of clarity in the causality between actions and impacts (difficulty in attributing causality), • long timescales over which impacts might occur, • poor availability of information and monitoring data relating to impacts

Table A2.7: Types of evaluation use

Instrumental	Conceptual	Strategic	Process-related
Evidence has a direct impact on policy	Evidence influences how stakeholders think about a policy area/issue	Evidence used for accountability and defending/promoting policy	Improved working processes in some way

Appendix 3 (Projects 1-23)

Project 01- Low Cost Resilience

Project Number	01
Project	***Supporting the Uptake of Low Cost Resilience for Properties at Risk of Flooding***
Description	The project's aim is to identify barriers and propose solutions to promote low cost flood approaches that would make properties at flood risk more resilient to damage from flood waters. This supports the long-term goal of enabling individuals and communities to take more ownership for the management of their flood risk and to recover more quickly as a result. The project used an action research approach, working closely with the at risk community to: • Update existing knowledge on sensible and low cost ways to make changes to properties during reinstatement and at other times when changes are happening as a result of renewal and transfer of properties. • Work with households, small businesses and their property advisers to understand how and when the adaptations can happen and what new training and advice is needed. • Design strategies to exploit opportunities to increase take up of measures with the support of householders, business owners and property professionals among others. • Work with the professionals, the community and other stakeholders to trial these strategies in one community. CEP led the work package on the trial (demonstration) evaluation and feedback to ensure that all the lessons were learned and shared as widely as possible. The team also attended events and organised evaluation activities. CEP supported the other teams, on a consultative basis, in designing the small scale trial and learning events as required. The project team was led by the University of the West of England (UWE) Centre for Floods Communities and Resilience and included CEP, MDA Associates and Birmingham City University (BCU).
Category	National policy intervention
Evaluation objectives	1. To evaluate the innovations and related materials to be developed under the project, and the forming of a Learning and Action Alliance (LAA) 2. To evaluate the demonstration phase (set up, during and after)
Methods used	Literature review, Observation, Interviews, Workshops/Events
Complexity	**Policy/Response-related:** • Degree of flexibility or tailoring/changes in the policy during implementation
	Impact-related: • Lack of clarity in the causality between actions and impacts • Availability of information and monitoring data relating to impacts
Evaluation use	• Conceptual use: evidence influence how stakeholders think about policy area/issue • Strategic use: evidence used for accountability and defending/promoting policy
Year	2015 - 2016
Scale	Local
Budget	<£20,000
Commissioning body	Department for Environment, Food and Rural Affairs (Defra)

Project 02 - Climate Change Strategy for Wales

Project Number	02
Project	**Evaluation of the Climate Change Strategy for Wales**
Description	Text Production of the Climate Change Commission for Wales' First Annual Report 2011 (with CAG Consultants). In the first year of reporting on the progress against the Climate Change Strategy delivery plans, progress was assessed using a review of action against the Delivery Plans, together with comments on the extent to which the delivery plans themselves are sufficiently ambitious. An evaluation framework was developed to perform two tasks: provide an overview evaluation framework for Climate Change Commission (the Commission) annual reviews for the next 5 years; and within this context, to provide a more detailed evaluation framework for the Commission's first annual review (2011). The overarching evaluation framework will provide a steer for the evaluation of Commission annual reviews for the next 5 years. The evaluation reviewed assessed the linked issues of: • The role and effectiveness of the Commission. • Delivery of the Welsh Government's (WG) commitments in the Climate Change Strategy for Wales (CCSW) and the measures set out in Emission Reduction and Adaptation Delivery Plans, with comment by the Commission on this delivery. • Delivery across wider sectors against the specific themes within the CCSW (with CAG)
Category	National policy intervention
Evaluation objectives	The first annual review focused on delivery against the Climate Change Strategy for Wales and Adaptation Delivery Plan and Delivery Plan for Emissions Reduction which were published in October 2010, and stakeholder reaction to the Climate Change Engagement Strategy published in November 2011, as well as the Commission's role and the effectiveness of its work programme. This assessment was made on two levels: • An assessment of the current progress against the adaptation and emission reduction delivery plans for the Climate Change Strategy which focuses primarily on the actions of the Welsh and UK Governments. • A wider assessment of progress in Wales within specific sectors, which explores the Welsh and UK Governments roles as well as those of wider sector groups and explores action and makes recommendations which sit within the context of the Climate Change Strategy but which often go beyond the scope of the current delivery plans themselves.
Methods used	Literature review, Surveys/Questionnaires, Interviews
Complexity	**Issue-related:** • Problem has multiple elements • Level of unpredictability in the problem **Policy/Response-related:** • Degree of flexibility or tailoring/changes in the policy during implementation **Impact-related:** • Availability of information and monitoring data relating to impacts
Evaluation use	• Instrumental use: evidence had a direct impact on policy • Conceptual use: evidence influence how stakeholders think about policy area/issue
Year	2011 - 2012
Scale	National
Budget	<£20,000
Commissioning body	Climate Change Commission for Wales
Weblink to report	http://www.cewales.org.uk/files/5314/4353/4874/Climate-Change-Commission-for-Wales-Document.pdf

Project 03 – Bioenergy public dialogue

Project Number	03
Project	Evaluation of BBSRC's Bioenergy public dialogue project
Description	CEP undertook an evaluation of the bioenergy public dialogue project, including an assessment of the effectiveness and value of the process, its impact and success. The Biotechnology and Biological Sciences Research Council's (BBSRC) Bioenergy public dialogue project was intended to allow a range of UK residents to clearly articulate their diverse perspectives on bioenergy so that these views, concerns and aspirations could inform future research and policies on bioenergy. The approach used a dispersed dialogue model, involving BBSRC researchers and other stakeholders organising dialogue events based on a common set of materials.
Category	National policy intervention
Evaluation objectives	• Provide an independent, unbiased evaluation of the project, including assessment of the effectiveness and value of the process, its impact and success • Help BBSRC further define the original aim, objectives and expected outcomes/outputs of the project to enable continuing effective evaluation • Provide information on developing best practice in public dialogue projects that can both inform the dialogue methodology as it progresses and be used in the future.
Methods used	Observation, Surveys/Questionnaires, Interviews, Steering group meetings/Expert advice
Complexity	Issue-related: • Problem has multiple elements
	Policy/Response-related: • Multiple components/elements included in the policy/programme/initiative
	Impact-related: • Availability of information and monitoring data relating to impacts
Evaluation use	Instrumental use: evidence had a direct impact on policy
Year	2012 - 2014
Scale	National
Budget	<£20,000
Commissioning body	Biotechnology and Bioscience Research Council (BBSRC)
Weblink to report	http://www.bbsrc.ac.uk/documents/bioenergy-dialogue-interim-report-governance-pdf/

Project 04 – Effectiveness of SEA Directive

Project Number	04
Project	**Study concerning the preparation of the report on the application and effectiveness of the SEA Directive (Directive 2001/42/EC)**
Description	The overall aim of the study contract is to provide information about the application and effectiveness of the SEA Directive for the period 2007-2014. This entails compiling, assessing and synthesising relevant information and evidence on the status of application and effectiveness of the SEA Directive across all 28 EU Member States, with due account taken of the respective national laws. The findings will be used by the Commission in preparing the second report on the application and effectiveness of the SEA Directive, due in 2016. The report will cover the period between July 2007 and July 2014. Furthermore, the findings of the implementation report will feed into any subsequent Regulatory Fitness and Performance Programme (REFIT) evaluation. A REFIT evaluation would constitute a more comprehensive assessment of: (i) effectiveness; (ii) efficiency; (iii) coherence; (iv) relevance, and (v) EU value-added. (led by Milieu) Services Provided: • Part of core team as well as providing expert adviser. • Input includes: reviewing scoping sheets at key milestones; advising on targets and methods for consultation; participating in webinar for peer review; assisting with final quality assurance; undertaking and advising on literature and evidence review; and peer review.
Category	EU policy intervention
Evaluation objectives	• provide the Commission with information on Member States' progress and challenges experienced in the application of the SEA Directive for the period 2007-2014. • contribute to the understanding of the extent to which Member States are implementing the SEA Directive (full implementation by 2020)
Methods used	Literature review, Surveys/Questionnaires, Interviews, Workshops/Events, Steering group meetings/Expert advice
Complexity	**Issue-related:** • Variability in the physical/environmental characteristics of the area/location
	Policy/Response-related: • Multiple agencies/actors/stakeholders involved or targeted by the policy • Competing/ interacting policies (at a UK or EU level)
Evaluation use	Instrumental use
Year	2015-2016
Scale	National/Multinational
Budget	<£20,000
Commissioning body	European Commission (DG Environment)
Weblink to report	**http://ec.europa.eu/environment/eia/pdf/study_SEA_directive.pdf**

Project 05 – Evaluation of pyrotechnic Directive

Project Number	05
Project	**Ex-post evaluation of the implementation by Member States of Directive 2007/23/EC on pyrotechnic articles**
Description	Led by Milieu. Economic expert opinion, including development of methodology and ex post assessment of economic impacts of implementation especially impacts on the internal market.
Category	EU policy intervention
Evaluation objectives	• Provide the Commission with an overview of the transposition and implementation and its main impacts on the free movement of pyrotechnic articles within the internal market.
Methods used	Literature review, Data/indicator review, Surveys/Questionnaires, Interviews, Steering group meetings/Expert advice, Develop case studies
Complexity	**Policy/Response-related:** • Multiple agencies/actors/stakeholders involved or targeted by the policy • Degree of flexibility or tailoring/changes in the policy during implementation
	Impact-related: • Availability of information and monitoring data relating to impacts
Evaluation use	• Instrumental use: evidence had a direct impact on policy
Year	2010-2011
Scale	National/Multinational
Budget	<£20,000
Commissioning body	European Commission

Project 06 – EC biocidal products impact assessment

Project Number	06
Project	**Assessing the impact of the revision of Directive 98/8/EC concerning the placing of biocidal products on the market**
Description	Part of the team of consultants carrying out a partial regulatory impact assessment of proposed options for the revision of the Directive, (with RPA Ltd., Milieu Ltd., and Hydrotox GmbH)
Category	EU policy intervention
Evaluation objectives	• support the Commission in the preparation of the proposal for the revision of the Directive 98/8/EC Concerning the Placing of Biocidal Products on the Market; • identify and evaluate the direct and indirect, positive and negative impacts related to the range of policy options available for addressing the current problems and shortcomings of the Directive on different stakeholders, including industry and in particular the small and medium-sized enterprises, the Member State Competent Authorities, the formulators and final users of biocidal products and on the environment
Methods used	Data/indicator review, Interviews, Workshops/Events, Steering group meetings/Expert advice
Complexity	**Policy/Response-related:** • Multiple agencies/actors/stakeholders involved or targeted by the policy • Degree of flexibility or tailoring/changes in the policy during implementation
	Impact-related: • Availability of information and monitoring data relating to impacts
Evaluation use	• Instrumental use: evidence had a direct impact on policy
Year	2007-2008
Scale	National/Multinational
Budget	£51,000 - 99,999
Commissioning body	European Commission (DG Environment)

Project 07 - Biodiversity Offsetting

Project Number	07
Project	***Evaluation of the Biodiversity Offsetting Pilot Phase***
Description	CEP undertook a two-year evaluation of the biodiversity offsetting pilot phase. Biodiversity offsetting is defined as conservation activities which are designed to deliver biodiversity benefits in compensation for losses, typically caused by development on a particular site, in a measurable way. Defra commissioned the project to evaluate the results of the six offsetting pilots and complementary projects. Broadly the evaluation considered how biodiversity offsetting can: • Help to use resources more effectively to deliver greater benefits for biodiversity. • Streamline the process of agreeing compensation for biodiversity loss as required by planning policy, in a cost effective way. At the same time the evaluation considered the pilots in relation to Defra's core principles of offsetting: not change existing levels of protection for biodiversity; deliver real benefits for biodiversity; be managed at the local level as far as possible; be as simple and straightforward as possible, for developers, local authorities and others; be transparent, giving clarity on how the offset calculations are derived and allowing people to see how offset resources are being used; and be good value for money. The project also considered whether biodiversity offsets deliver benefits as compared to the existing approach (i.e. a case-by-case discussion of the mitigation or compensation required by planning policy). The final project report included findings under five main themes (i.e. governance; process and management; legal and development planning; costs; and ecological implementation and monitoring) and made recommendations to Defra in light of the findings of the evaluation of the offsetting pilots and complementary projects and in terms of whether and how offsetting can help make better use of resources for delivering biodiversity benefits.
Category	Programme level policy intervention
Evaluation objectives	1. Estimate the existing costs of compensation for residual biodiversity loss 2. Reveal the existing costs of compensation for residual biodiversity loss, including delays caused by negotiations/surveys, and issues around any land undeveloped because of on-site compensation. 3. Establish the relative frequency of occurrence for these costs and provide information in a way useable in an impact assessment, noting the risks and sensitivities imposed by the research methodology.
Methods used	Literature review, Data/indicator review, Interviews, Workshops/Events, Steering group meetings/Expert advice, Develop case studies
Complexity	**Policy/Response-related:** • Multiple agencies/actors/stakeholders involved or targeted by the policy • Degree of flexibility or tailoring/changes in the policy during implementation **Impact-related:** • Timescales over which impacts might occur
Evaluation use	• Instrumental use: evidence had a direct impact on policy • Conceptual use: evidence influence how stakeholders think about policy area/issue
Year	2012-2014
Scale	Local/National
Budget	£200,000 - £300,000
Commissioning body	Department for the Environment, Food and Rural Affairs (Defra)
Weblink to report	http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=18229

Project 08 – Ex-ante Evaluation and SEA of Wales RDP

Project Number	08
Project	Ex-Ante Evaluation and Strategic Environmental Assessment (SEA) of the Wales Rural Development Plan (2007-2013)
Description	The Wales Rural Development Plan (RDP) for the period 2007-13 offered a new strategic approach to rural development focusing on three core objectives: (1) increasing the competitiveness of the agricultural and forestry sector through support for restructuring; (2) enhancing the environment and countryside through support for land management; and (3) enhancing the quality of life in rural areas and promoting the diversification of economic activities through measures targeting the farm sector and other rural actors. The ex-ante evaluation took place between October, 2005 and September, 2006 and was conducted by Agra CEAS Consulting Ltd. The Strategic Environmental Assessment (SEA) was undertaken by Collingwood Environmental Planning alongside the development of the Wales RDP seeking to ensure that the RDP contributes positively to the high level of environmental protection expected of EU funding programmes. The SEA aimed to ensure that likely significant effects on the environment of implementing the RDP, and of reasonable alternatives, were identified, described, evaluated and taken into account before the plan was adopted. For the SEA a causal change analysis was used as a method for understanding pathways to impact.
Category	Programme level policy intervention
Evaluation objectives	The objective of the ex-ante evaluation was to support the preparation of the Wales Rural Development Plan proposals, gather information and carry out analyses to help to ensure that the delivery of the RDP objectives will be successful and that reliable evaluation will be subsequently possible. The overall aims of the SEA were to: • provide for a high level of environmental protection; • ensure that likely significant effects on the environment of implementing the RDP are identified, described, evaluated and taken into account before the plan is adopted; and that, • reasonable alternatives, taking into account the objectives and geographical scope of the plan, are evaluated for their likely significant effects and inform the nature and content of the proposed RDP.
Methods used	Literature review, Data/indicator review, Workshops/Events, Steering group meetings/Expert advice, Develop case studies
Complexity	Issue-related: • Problem has multiple elements • Variability in the physical/environmental characteristics of the area/location Policy/Response-related: • Degree of flexibility or tailoring/changes in the policy during implementation
Evaluation use	Strategic use: evidence used for accountability and defending/promoting policy
Year	2005 - 2006
Scale	National
Budget	£21,000 - 50,000
Commissioning body	Welsh Assembly
Weblink to report	Ex-ante evaluation: http://www.ceasc.com/images/content/2239%20final%20report.pdf SEA: http://www.ceasc.com/Images/Content/2239%20SEA.pdf

Project 09 - Scottish Pathfinder SEA

Project Number	09
Project	Scottish Government Strategic Environmental Assessment (SEA) Pathfinder Research Project
Description	Commissioned along with the EnviroCentre by the Scottish Executive to a three year research project to evaluation current practice in SEA in collaboration with to eight SEA case studies across Scotland with a view to help inform the development of good practice in SEA in Scotland.
Category	Programme level policy intervention
Evaluation objectives	• Develop the evaluation method for assessing case study SEAs; • Develop evaluation pro-formas for on-going assessment and monitoring of the case study SEAs; • Examine and evaluate current practice in collaboration with SEA case studies and the project steering group; • Review current SEA practice to identify good practice; • Identify recommendations for good practice in collaboration with SEA case studies and the project steering group; and • Propose recommendations for action to implement good practice in preparation for Stage 2.
Methods used	Literature review, Data/indicator review, Observation, Interviews, Steering group meetings/Expert advice
Complexity	Issue-related: • Problem has multiple elements Policy/Response-related: • Multiple agencies/actors/stakeholders involved or targeted by the policy • Degree of flexibility or tailoring/changes in the policy during implementation
Evaluation use	• Instrumental use: evidence had a direct impact on policy • Strategic use: evidence used for accountability and defending/promoting policy
Year	2005 - 2008
Scale	Local/National
Budget	£51,000 - 99,999
Commissioning body	Scottish Government

Project 10 – Flood Awareness Wales

Project Number	10
Project	Flood Awareness Wales Community Engagement Review
Description	CEP undertook an independent review of Natural Resources Wales' 'Flood Awareness Wales Community Engagement Programme' (FAW), covering the period 2010 - 2015. The main aims of the evaluation were to assess the effectiveness of current approaches to community engagement; and to review local, national and international best practice in order to provide evidence and recommendations for future practice to increase community flood resilience in Wales. A simple logic model was developed to describe how FAW's resources and activities were expected to produce the intended outcomes and impacts. Collecting and analysing evidence from each of the FAW activities made it possible to see how far assumptions were borne out and the way in which the different activities contributed to the programme's results.
Category	Programme level policy intervention
Evaluation objectives	1. To assess the effectiveness of NRW's approaches to date, specifically in relation to the current model of operation - development of flood plans, which are supported and maintained by local flood volunteers 2. To provide evidence and recommendations drawing on local, national and international best practice and approaches to date, to inform future practice in increasing community flood resilience across Wales
Methods used	Literature review, Surveys/Questionnaires, Interviews
Complexity	Issue-related: • Sensitivity to socio-demographic characteristics of the area/target population Policy/Response-related: • Degree of flexibility or tailoring/changes in the policy during implementation Impact-related: • Lack of clarity in the causality between actions and impacts
Evaluation use	• Instrumental use: evidence had a direct impact on policy • Conceptual use: evidence influence how stakeholders think about policy area/issue • Strategic use: evidence used for accountability and defending/promoting policy
Year	2015 - 2016
Scale	Local/Regional/National
Budget	£21,000 - 50,000
Commissioning body	Natural Resources Wales
Weblink to report	https://naturalresources.wales/media/679872/20160706-independent-faw-review-executive-summary-english.pdf

Project 11 – Ex-ante and SEA of the Scottish RDP

Project Number	11
Project	Ex-Ante Evaluation and Strategic Environmental Assessment of the proposed Scottish Rural Development Programme (SRDP) 2014-2020
Description	Undertook the Strategic Environmental Assessment (SEA) of the proposed Scotland Rural Development Programme (SRDP) 2014-2020 (with Agra CEAS Consulting Ltd and SRUC). CEP undertook the overall project management and coordination of the SEA, developed the approach and undertook all aspects of the SEA, including engagement with stakeholders.
Category	Programme level policy intervention
Evaluation objectives	The SEA Objective was to assess four broad strategic alternatives to the SRDP proposal.
Methods used	Literature review, Data/indicator review, Workshops/Events, Steering group meetings/Expert advice
Complexity	Issue-related: • Problem has multiple elements • Variability in the physical/environmental characteristics of the area/location
	Impact-related: • Multiple types / range of possible/expected outcomes and impacts
Evaluation use	• Instrumental use: evidence had a direct impact on policy • Strategic use: evidence used for accountability and defending/promoting policy
Year	2012 - 2014
Scale	National
Budget	£21,000 - 50,000
Commissioning body	Scottish Government
Weblink to report	http://www.gov.scot/Topics/farmingrural/SRDP/SRDP20142012/SRDP20142020ExAnteEvaluationSEA

Project 12 – Nature Improvement Areas

Project Number	12
Project	**Monitoring and Evaluation of Nature Improvement Areas: Phase 2**
Description	Nature Improvement Areas (NIAs) were places where a shared vision for the development of the natural environment existed among a wide partnership of local people, including statutory and voluntary sectors. They were places in which real improvements to the local environment could be achieved over large areas by enlarging and enhancing existing wildlife sites, improving ecological connectivity and creating new sites. An initial 12 NIAs started work in April 2012, with governmental support through the NIA grant scheme. These were partnerships of local authorities, local communities and landowners, the private sector and conservation organisations, promoting opportunities for restoring and connecting nature on a significant scale. The NIA Monitoring and Evaluation (M&E) Phase 2 project was a three year project (February 2013 - November 2015) which built on the outcomes of the Phase 1 Scoping Study - Developing a framework for design, monitoring and evaluating NIAs. The project gathered quantitative and qualitative evidence and assessed the progress and achievements of the NIAs over the three year grant funded period, as well as learning from the NIA initiative to inform future integrated natural environment initiatives. Support to the NIAs' monitoring and evaluation activities included: research into and developing approaches to measure social and economic impacts and wellbeing benefits of the NIAs; developed a case study based approach to enable NIAs to report on their contribution to local social and economic outcomes (e.g. contributions to local employment, mobilisation and training of volunteers etc.); developed guidance for NIAs to help them develop local community and visitor surveys / questionnaires; and developing approaches related to monitoring and evaluating ecosystem services and habitat connectivity. Other elements of the project included: research to test and increase understanding of different approaches to assess the difference the NIAs made over and above what would have happened anyway (the counterfactual); and a scoping study to design the monitoring and evaluation of the Countryside Stewardship facilitation fund (CSFF).
Category	Programme level policy intervention
Evaluation objectives	The overall aims of the Phase 2 project were to: • gather evidence of approaches used within the NIAs and their outcomes, to maximise learning from the pilots and build a practical evidence base to inform future extension of the NIA approach; and, • assess the individual and aggregated contribution of the NIA pilots towards meeting the outcomes included in the Natural Environment White Paper and other agreed policy outcomes.
Methods used	Data/indicator review, Surveys/Questionnaires, Interviews, Workshops/Events, Steering group meetings/Expert advice, Develop case studies
Complexity	**Policy/Response-related:** • Multiple components/elements included in the policy/programme/initiative **Impact-related:** • Timescales over which impacts might occur • Availability of information and monitoring data relating to impacts
Evaluation use	• Conceptual use: evidence influence how stakeholders think about policy area/issue • Strategic use: evidence used for accountability and defending/promoting policy • Process-related use: improved working processes in some way
Year	2023 - 2015
Scale	Local/National
Budget	>£300,000
Commissioning body	Department for Environment, Food and Rural Affairs (Defra)
Weblink to report	http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&ProjectID=18555&FromSearch=Y&Publisher=1&SearchText=WC1061&SortString=ProjectCode&SortOrder=Asc&Paging=10#Description

Project 13 – Enhancing ex-post evaluation of FCRM

Project Number	13
Project	**Enhancing ex-post evaluation of flood and coastal erosion risk management plans and schemes**
Description	This project investigated the practice of, and barriers to, ex-post evaluation of flood and coastal erosion risk management Strategy Plans and schemes and considered how evaluation practice could be enhanced, including through guidance to Risk Management Authority practitioners. The project had two phases: Phase 1 reviewed and collated information from documents, literature and interviews with key stakeholders in order to establish what constitutes good practice ex-post evaluation, how current practice differs and any reasons for this and to propose options for improving practice. Phase 2 explored the potential for a 'light touch' approach to ex-post evaluation using documentary analysis and interviews in two case studies. CEP led a team which included Middlesex University's Flood Hazard Research Centre and HR Wallingford.
Category	Programme level policy intervention
Evaluation objectives	Phase 1 Objectives: • To review evaluation practice in respect of strategies and schemes and how this compares with good practice. • To highlight individual examples of good practice. • To identify existing barriers to good evaluation practice, including institutional and cultural issues, adequacy of guidance, and those to do with methodological and evidence challenges, including adequacy of monitoring arrangements and data collection. • To consider how evaluation practice could realistically be improved within existing institutional arrangements, and taking due account of proportionality. • To consider the extent to which evaluation practice could be enhanced through updated guidance to practitioners Phase 2 objectives: • To develop outline methods and approaches for undertaking EPE • To explore the potential for 'light touch' approaches to EPE • To prepare a shortlist of four schemes suitable for case studies on EPE and take forward two of them as full case studies using the methods developed for EPE • To investigate how existing data and approaches could be used within EPE and what new data or approaches might be needed going forward • To reflect on lessons arising from EPE and consider how learning can be taken forward.
Methods used	Literature review, Interviews, Workshops/Events, Steering group meetings/Expert advice, Develop case studies
Complexity	**Policy/Response-related:** • Multiple agencies/actors/stakeholders involved or targeted by the policy
	Impact-related: • Timescales over which impacts might occur • Availability of information and monitoring data relating to impacts
Evaluation use	• Conceptual use: evidence influence how stakeholders think about policy area/issue • Process-related use: improved working processes in some way
Year	2014 - 2015
Scale	Local/National
Budget	£51,000 - 99,999
Commissioning body	Department for Environment, Food and Rural Affairs (Defra)

Project 14 – Land Use Strategy: Delivery Evaluation

Project Number	14
Project	**Land Use Strategy: Delivery Evaluation Project**
Description	This project aimed to provide evidence to Ministers on the effectiveness of the approach set out in the LUS with regard to delivering the Strategy's Principles. It evaluated current land use delivery mechanisms, to ascertain their effectiveness in translating the Principles of Sustainable Land Use into decision making on the ground, including how stakeholders and communities are being involved in land use decisions. The approach involved longitudinal qualitative assessment of 11 case study projects from multiple sectors across Scotland. (with the University of Strathclyde)
Category	Programme level policy intervention
Evaluation objectives	The detailed objectives were: • To assess each process/approach in terms of how well it is able (implicitly or explicitly) to translate the high level LUS objectives into decision making on the ground • To identify where and how the Principles of the LUS are successfully being applied; to investigate why methods are working well and identify successful aspects which might be applied more generally across Scotland in a range of different circumstances • To identify any barriers to the application of the LUS Principles, why this is the case and what lessons can be learned for more general application across Scotland • To use the evidence gathered across the range of projects to highlight emerging themes on how best to apply the Principles for Sustainable Land Use to different circumstances and processes across Scotland. Where possible this should focus on messages that will be useful in specific circumstances, and for a range of groups of decision makers and stakeholders
Methods used	Literature review, Interviews, Workshops/Events
Complexity	**Issue-related:** • Variability in the physical/environmental characteristics of the area/location
	Policy/Response-related: • Multiple agencies/actors/stakeholders involved or targeted by the policy
	Impact-related: • Interactions between components of a policy
Evaluation use	• Instrumental use: evidence had a direct impact on policy • Conceptual use: evidence influence how stakeholders think about policy area/issue • Strategic use: evidence used for accountability and defending/promoting policy • Process-related use: improved working processes in some way
Year	2012 - 2014
Scale	Regional/National
Budget	£21,000 - 50,000
Commissioning body	Scottish Government
Weblink to report	http://www.gov.scot/Publications/2014/05/5854

Project 15 – LUS Forestry Pilot Studies

Project Number	15
Project	**Evaluation of the Land Use Strategy (LUS) Forestry Focused Sub-Regional Pilot Studies**
Description	Recognising the diverse range of important benefits that well-designed, well-managed forests can bring, Scotland has an ambitious target to create 100,000ha of new woodland during the period 2012-2022. However, Scotland's land resource is under pressure to deliver a range of benefits and conflicts can occur between competing land uses (e.g. farming and forestry). Responding to these pressures, the Woodland Expansion Advisory Group (WEAG) for Scotland published its report in 2012 advising on the types of land that are best suited for tree planting in Scotland, in the context of other land uses. The WEAG also recognised that current regional level forest planning may not be the most suitable approach for identifying woodland expansion opportunities, particularly in constrained and / or highly heterogeneous landscapes. Consequently, the WEAG advised that sub-regional forest planning pilots should be undertaken across Scotland with a view to rolling out the approach more widely in the future, where tangible benefits could be identified. The purpose of this project therefore was to evaluate the three WEAG sub-regional forest planning pilots to inform policy decisions regarding the future use of this approach in Scotland. The evaluation objectives considered: the processes that helped to shape each pilot; the potential of the pilots to help deliver multifunctional forestry in line with Scotland's Land Use Strategy (LUS); the balance of costs and benefits for each pilot; and the strengths, weaknesses and good-practice inherent to the pilots. The evaluation involved extensive review of documentary evidence, in-depth semi-structured interviews with a range of stakeholders associated with the pilots (e.g. Forestry Commission Scotland, environmental regulators, local authorities, NGOs, private sector) and criteria based evaluation.
Category	Programme level policy intervention
Evaluation objectives	1. To examine and evidence the processes that helped shape each pilot project 2. To assess and examine each pilot in terms of its potential ability to guide local level land use decision-making to help meet the overall objectives of the LUS 3. To provide evidence on whether the pilot projects have influenced local level land use decision-making to help meet the overall objectives of the LUS 4. To provide evidence and views from responsible organisations and stakeholders in the pilot areas on whether the benefits of the pilots justify the costs and resources required for their development 5. To identify strengths, weaknesses and good-practice in the pilot frameworks
Methods used	Literature review, Interviews
Complexity	**Issue-related:** • Variability in the physical/environmental characteristics of the area/location
	Policy/Response-related: • Multiple agencies/actors/stakeholders involved or targeted by the policy
	Impact-related: • Multiple types / range of possible/expected outcomes and impacts
Evaluation use	• Conceptual use: evidence influence how stakeholders think about policy area/issue • Process-related use: improved working processes in some way
Year	2014 - 2015
Scale	Regional/National
Budget	<£20,000
Commissioning body	Forestry Commission Scotland
Weblink to report	http://scotland.forestry.gov.uk/images/evaluation-lus-final-report.pdf

Project 16 – Community Flood Forum Evaluation

Project Number	16
Project	**Evaluation of the Communities Prepared project**
Description	The Communities Prepared project is supporting and training people in local communities and encouraging them to volunteer and respond to flooding and other emergencies. Key to this new programme is a training toolkit to help all communities become more resilient to challenges that they may face. Although focussing on flooding, this programme will help communities prepare for a range of potential emergencies. In the first phase of the project eight communities in South West England and two in the North will be supported. Once this pilot has been completed, the Communities Prepared Partnership would look to support communities across the country to become more resilient and better prepared. CEP is carrying out the evaluation of the project. This involves developing indicators community and household resilience to emergencies, establishing a baseline against which to measure change over the period of the project and designing a suite of evaluation tools. (led by Groundwork South, with Exeter University and Cornwall Community Flood Forum) Services provided: • Design and application of evaluation framework and tools. • Advice on project design, e.g. selection of communities.
Category	Programme level initiative
Evaluation objectives	• Test the effectiveness of the Train the Trainer approach in developing community volunteers' confidence and ability to remain safe and play an effective role before, during and after an emergency such as a flood. • Test the ability of the Train the Trainer approach in delivering consistently delivered training and learning. • Identify how and in what ways the development of the confidence and capacities of community volunteers contributes to making individuals within the communities involved better informed and able to prepare for and respond safely to emergencies such as flooding. • Increase the confidence and ability to remain safe of community volunteers and increase their capacity to play an effective role before, during and after an emergency. • Ensure that members of the community are better prepared for, and more resilient to, the impact of emergencies.
Methods used	Data/indicator review, Observation, Surveys/Questionnaires, Interviews, Steering group meetings/Expert advice,
Complexity	**Issue-related:** • Problem has multiple elements • Sensitivity to socio-demographic characteristics of the area/target population
	Policy/Response-related: • Multiple agencies/actors/stakeholders involved or targeted by the policy
Evaluation use	Evaluation still ongoing
Year	2015-2017
Scale	Local/Regional/National
Budget	£21,000 - 50,000
Commissioning body	Big Lottery Fund

Project 17 – Climate Mitigation through ECF

Project Number	17
Project	**Childrens Investment Fund for the Future (CIFF) Evaluation of the European Climate Foundation (ECF)**
Description	Project to evaluate the effectiveness of the European Climate Foundation (ECF) in meeting its programme objectives (over the funding period 2011 – 2015). The objectives of this impact evaluation were: to help guide the Children's Investment Fund for the Future (CIFF) Climate Change Advisory Board in taking strategic decisions on support to climate change action in Europe and future grants to ECF; and to provide insights to guide CIFF's strategic approach to climate change in Europe. The evaluation sought to assess progress towards reducing emissions from the energy sector (and, where relevant, economy-wide) in Europe and the contribution of ECF to that reduction. In doing so in-depth theories of change developed by ECF for each of their interventions were explored and tested through mixed methods including: document review; semi-structured interviews; online surveys tailored for different stakeholder groups; observation of ECF planning meetings; and a workshop. (project led by URSUS Consulting) Services provided: CEP supported the evaluation by providing: • Expert input to the evaluation design, including advice on indicators, quantification and theory of change. • Leading on the development of options for attribution of change (e.g. changes in policy, energy mix, emissions) to the activities of ECF and its network of partners. • Undertaking semi-structured interviews with stakeholders in ECF partner organisations and the Commission. • Drafting final report text on attribution.
Category	Programme level initiative
Evaluation objectives	Help guide CIFF's Climate Change Advisory Board (CCAB) in taking strategic decisions on support to climate change action in Europe and future grants to ECF.
Methods used	Literature review, Data/indicator review, Observation, Surveys/Questionnaires, Interviews, Workshops/Events, Steering group meetings/Expert advice
Complexity	**Policy/Response-related:** • Multiple components/elements included in the policy/programme/initiative • Multiple agencies/actors/stakeholders involved or targeted by the policy
	Impact-related: • Lack of clarity in the causality between actions and impacts
Evaluation use	• Strategic use: evidence used for accountability and defending/promoting policy
Year	2015
Scale	National/Multinational
Budget	£51,000-99,999
Commissioning body	Childrens Investment Fund for the Future (CIFF)
Weblink to report	No formal publication expected

Project 18 – Defra CaBA Phase 2 Evaluation

Project Number	18
Project	**Catchment Base Approach (CaBA): Monitoring and evaluation (Phase 2) and wider adoption of CaBA for the period 2013-15**
Description	The Catchment Based Approach (CaBA) embeds collaborative working at a river catchment scale to deliver cross cutting improvements to water environments. Community partnerships, bringing local knowledge and expertise, are active in over 100 Water Framework Directive catchments in England. The purpose of the project was to evaluate the activities underpinning wider adoption of the Catchment Based Approach. This involved assessing the effectiveness of national funding and support structures and looking at how effective CaBA partnerships are in supporting community / public engagement. The project reviewed how CaBA is aligning with River Basin Management Planning and gathered evidence on good practice, lessons learned and additional support needs. As an outcome CEP developed a suite of self-assessment tools for catchment based approach (CaBA) partnerships to enable them to quickly assess how well they are performing in terms of recognised milestones and good practice from around the country. (Led by Cascade Consulting, with eftec and Lancaster University). Services provided: • Technical support on the development of the evaluation framework. • Survey design and content. • Evaluation reporting and policy recommendations. • Design and testing of self-assessment tools.
Category	Programme level initiative
Evaluation objectives	To evaluate the support and activities underpinning the wider adoption of the Catchment Based Approach ('CaBA').
Methods used	Surveys/Questionnaires, Interviews
Complexity	**Issue-related:** • Problem has multiple elements **Policy/Response-related:** • Multiple components/elements included in the policy/programme/initiative • Multiple agencies/actors/stakeholders involved or targeted by the policy
Evaluation use	• Strategic use: evidence used for accountability and defending/promoting policy • Process-related use: improved working processes in some way
Year	2014-2015
Scale	Local/National
Budget	£100,000-199,999
Commissioning body	Department for Environment, Food and Rural Affairs (Defra)
Weblink to report	http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=19337

Project 19 – CEFAS MCCIP mid-term review

Project Number	19
Project	**Independent Mid-Term review of the Marine Climate Change Impacts Partnership (MCCIP) Work Programme**
Description	The Marine Climate Change Impacts Partnership (MCCIP) brings together many of the UK's key organisations with marine interests and responsibilities, from central Government, the devolved administrations, advisory and regulatory agencies, the scientific community and NGOs. The principal aim of MCCIP is to provide a co-ordinating framework for the UK, so as to be able to transfer high quality evidence on marine climate change impacts, and guidance on adaptation and related advice, to policy advisors and decision-makers. This project assessed progress of the Phase II MCCIP work programme (2010-2015) through a mid-term review. Services Provided: • Independent evaluation using data from online questionnaire and in-depth interviews with a selection of MCCIP Steering group members, direct beneficiaries, and the MCCIP secretariat. • Assessment and synthesis of the progress to date in achieving MCCIP aim and objectives and delivery of MCCIP key 'outputs' and 'outcomes'.
Category	Programme level initiative
Evaluation objectives	Assess the benefits and value for money from MCCIP work perceived by its 'direct' and 'indirect' beneficiaries, as well as contributors to MCCIP work on the following: • Progress to date in achieving MCCIP aims and objectives; • Use of MCCIP products; • Progress in achieving the MCCIP interim outcomes for Phase II; • Overall value of the partnership to Steering Group members, direct beneficiaries and participants.
Methods used	Literature review, Surveys/Questionnaires, Interviews
Complexity	**Policy/Response-related:** • Multiple components/elements included in the policy/programme/initiative • Multiple agencies/actors/stakeholders involved or targeted by the policy
	Impact-related: • Lack of clarity in the causality between actions and impacts
Evaluation use	• Conceptual use: evidence influence how stakeholders think about policy area/issue • Strategic use: evidence used for accountability and defending/promoting policy
Year	2014
Scale	National
Budget	<£20,000
Commissioning body	Centre for Environment, Fisheries and Aquaculture Science (CEFAS)
Weblink to report	http://www.mccip.org.uk/media/1444/120514-mccip-mtr-final.pdf

Project 20 – Defra FRCP Evaluation

Project Number	20
Project	**Flood Resilience Community Pathfinder Evaluation**
Description	CEP was leading a consortium to evaluate the Community Flood Resilience Pathfinder Scheme (Defra) and ensure that lessons can be learnt. This includes developing a Rapid Evidence Assessment, designing the framework for the evaluation, providing advice and support for the evaluation plans being undertaken by the Pathfinders and implementing a scheme-level evaluation The evaluation follows UK Government good practice guidance - specifically the Magenta Book. It started with a Rapid Evidence Assessment (REA) to provide an overview and synthesis of the available evidence on the topic to inform the development of an evaluation framework. In parallel, the team worked with each of the pathfinder projects to understand their focus and priorities and any methods of evaluation they have put in place. The results of these two strands combined to generate evaluation questions and indicators on the key issues for community resilience to flooding such as identifying and targeting measures to more vulnerable groups, empowering community members and building social capital, strengthening financial resilience and stimulating a change to more sustainable behaviours in terms of managing flood risk and responding to flooding. The final evaluation identified where measurable change has been achieved in reducing flood risk and increasing community resilience since the beginning of the Pathfinder scheme. It pointed to lessons and good practice which could be applied in other communities. The CEP consortium includes: the Flood Hazard Research Centre at Middlesex University; the University of Surrey; Northumbria University; the Centre for Evidence and Policy at King's College London; and nef consulting Limited. Services Provided: • Overall project management and coordination. • CEP is involved in delivering all aspects of the project as described above.
Category	Programme level initiative
Evaluation objectives	The specific objectives of the evaluation are to: • Report on the progress that projects are making; • Provide advice and support for the projects' own evaluations to ensure that assumptions and results are consistent across the scheme; • Report on the scheme and its impact.
Methods used	Literature review, Data/indicator review, Surveys/Questionnaires, Interviews, Workshops/Events, Steering group meetings/Expert advice, Develop case studies
Complexity	**Issue-related:** • Problem has multiple elements **Impact-related:** • Lack of clarity in the causality between actions and impacts • Timescales over which impacts might occur
Evaluation use	• Conceptual use: evidence influence how stakeholders think about policy area/issue
Year	2013-2015
Scale	Local/National
Budget	£200,000 - £300,000
Commissioning body	Department for Environment, Food and Rural Affairs (Defra)
Weblink to report	http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=18744

Project 21 – Defra Evaluation of CaBA Pilots

Project Number	21
Project	**Evaluation of the catchment-based approach (CaBA) - pilot stage**
Description	The UK Government was undertaking a fundamental review of river basin planning (RBP) strategy in the context of the European Water Framework Directive (WFD), with a new focus on institutional arrangements and processes. This included a requirement for engagement of stakeholders in new ways in order to ensure common ownership of problems and their solutions, linked with a more local, catchment based approach to RBP implementation. There was also an emphasis on joining up the previously separate water and land management policy issues, using the common lens of ecosystem services to ensure accountability of the multiple benefits provided by the natural environment. A series of catchment-level partnerships were developed through a pilot phase (May 2011 – December 2012) to test these new approaches. The intention was that all these pilots "will establish the right level of spatial targeting to address sources of water pollution and explore the most effective ways to engage partners. The pilots also aimed to establish how best to achieve integrated, multiple environmental outcomes", using an adaptive management approach. The UK Government's aim was that the pilots provide a means for stakeholders to learn together how to develop effective, partnership-based approaches for catchment improvement. As a basis for this learning, Defra was keen to see the pilots test a range of approaches. The purpose of this evaluation was to assess whether these potentially significant returns can be achieved through a relatively modest investment in the new, catchment-based approach, and if so, which approaches work best, where, for whom and why. (Project led by Cascade Consulting) • CEP was a core team member, including involvement in the project management group. • CEP led on the learning strand of this evaluation identifying the key issues relating to the establishment and delivery of the catchment based approach and supported the pilots in learning how to address the issues, supported the national rollout of CaBA and providing guidance on good practices. This included evaluating a requirement for engagement of stakeholders in new ways in order to ensure common ownership of problems and their solutions. • CEP also produced a Guide to Collaborative Catchment Management (CaBA)
Category	Programme level initiative
Evaluation objectives	The key objectives of the evaluation were to: • identify the key issues relating to the establishment and delivery of the CaBA; • develop an understanding of the costs and benefits to support future policy recommendations; • support the wider adoption of the approach by detailing lessons learned
Methods used	Observation, Surveys/Questionnaires, Interviews, Workshops/Events, Develop case studies
Complexity	**Policy/Response-related:** • Degree of flexibility or tailoring/changes in the policy during implementation **Impact-related:** • Lack of clarity in the causality between actions and impacts • Availability of information and monitoring data relating to impacts
Evaluation use	• Instrumental use: evidence had a direct impact on policy • Conceptual use: evidence influence how stakeholders think about policy area/issue • Strategic use: evidence used for accountability and defending/promoting policy • Process-related use: improved working processes in some way
Year	2011-2013
Scale	Local/National
Budget	>£300,000
Commissioning body	Department for Environment, Food and Rural Affairs (Defra)
Weblink to report	http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=17943

Project 22 – Ex-post of Cohesion Policy Programmes

Project Number	22
Project	**Ex-post evaluation of Cohesion Policy Programmes 2000-2006 Co-Financed by the European Fund for Regional Development (Objective 1 and 2) Work Package 5b: Environment and Climate Change**
Description	The overall goal of this study was to evaluate the contribution of the European Regional Development Fund (ERDF) toward the implementation of EU environmental strategies between 2000 and 2006. During this programming period, 21% of the overall ERDF budget was allocated for environmental interventions, with ERDF environmental support in the EU-25 totalling €25.5 billion between 2000 and 2006. CEP's specific role was to provide the environmental expert for the UK as part of this ex post evaluation: writing the overview note for the United Kingdom. Collecting UK specific information to support the delivery of the project; undertaking a regional case study; and undertaking analysis of the findings and producing reports. (with ADE and Agrotec).
Category	Programme level initiative
Evaluation objectives	Evaluate the contribution of the European Regional Development Fund (ERDF) toward the implementation of EU environmental strategies between 2000 and 2006.
Methods used	Literature review, Data/indicator review, Observation, Interviews, Develop case studies
Complexity	**Policy/Response-related:** • Multiple components/elements included in the policy/programme/initiative • Geographic spread/ scale of the policy response
	Impact-related: • Lack of clarity in the causality between actions and impacts
Evaluation use	• Strategic use: evidence used for accountability and defending/promoting policy
Year	2008-2009
Scale	Local/Regional/Multinational
Budget	<£20,000
Commissioning body	European Commission (DG Regio)
Weblink to report	http://ec.europa.eu/regional_policy/sources/docgener/evaluation/expost2006/wp5b_en.htm

Project 23 – New Forest Pathfinder Project

Project Number	23
Project	**New Forest Pathfinder Project - evaluation of stakeholder participation and engagement processes**
Description	Led an evaluation of stakeholder participation and engagement processes in the New Forest. (with Shared Practice).
Category	Programme level initiative
Evaluation objectives	• review and evaluate existing stakeholder involvement in Forestry Commission consultation forums on land management in the New Forest. • review and evaluate methods that other key statutory agencies use to conduct stakeholder involvement in consultation for land management works in the New Forest. The objective being to identify any potential for agencies to work together to deliver stakeholder involvement.
Methods used	Literature review, Observation, Surveys/Questionnaires, Interviews, Workshops/Events, Develop case studies
Complexity	**Issue-related:** • Problem has multiple elements
	Policy/Response-related: • Multiple agencies/actors/stakeholders involved or targeted by the policy
	Impact-related: • Multiple types / range of possible/expected outcomes and impacts
Evaluation use	• Conceptual use: evidence influence how stakeholders think about policy area/issue
Year	2006-2007
Scale	Local
Budget	<£20,000
Commissioning body	Forestry Commission

Appendix 4 Evaluation Frameworks

- **Qualitative Evaluation (HM Treasury 2012).** Supplementary to the Magenta Framework the UK government has published a framework for assessing qualitative evaluations concerned with the development and implementation of social policy, programmes and practice. The document responds to widespread concerns about rigour and robustness of the guiding principles, data collection methodologies as well as analysis of evidence, arguing that qualitative research "*should be assessed on its own terms within premises that are central to its purpose, nature and conduct"* (HM Treasury, 2012). Building on previous guides? it develops a new framework with a particular focus on the methods used most extensively in government-based evaluations, namely, interviews, focus groups, observation and documentary analysis. Underpinning the framework are four guiding principles, each with a series of quality indicators that have been developed for assessment purposes. According to the framework a qualitative evaluation should be:
  - ***contributory*** *in advancing wider knowledge or understanding about policy, practice, theory or a particular substantive field;*
  - ***defensible*** *in design by providing a research strategy that can address the evaluative questions posed;*
  - ***rigorous*** *in conduct through the systematic and transparent collection, analysis and interpretation of qualitative data;*
  - ***credible*** *in claim through offering well-founded and plausible arguments about the significance of the evidence generated.*

- **Public dialogue (Sciencewise, 2016).** Building on *Sciensewise's Guiding Principles* (Sciencewise, 2013), a guide developed to support the UK government's approach to public dialogue, Sciencewise recently developed a framework for assessing the quality of public dialogue. A detailed look at the evaluation framework reveals factors that can be used to assess how successful the evaluation was. These are presented in the second column of Table A4.1, corresponding to the relevant evaluation activity in the first column.

Table A4.1 Sciencewise evaluation success factors

Clear scope for the evaluation	• **evaluation identified lessons emerging from the process and impacts** • **evaluation contributed to the design and delivery of the project**[5] • **timing - the evaluation captured the entire project from the early stages and throughout the public engagement** • **appropriateness of the evaluation design in assessing success against objectives, participants expectations, value for money, quality of engagement etc.**
Analytical frameworks and criteria	• the specific evaluation framework used was identified • evaluation used clear criteria for the assessment of effectiveness • any assumptions and hypothesis were recognised and discussed in terms of how they affected the design and output of the evaluation • any unexpected outcomes and consequences identified and discussed

[5] Applicable if formative evaluation

Evaluation reporting	• rationale for structure and form of the evaluation was discussed • the audience for the evaluation report identified • contributions of the evaluation to openness, transparency and accountability were discussed

Looking into practical examples of Sciencewise's evaluations, and despite those preceding the above methodological guidance, two documents provide an example of using theory of change as a framework to evaluate the Sciencewise programme*: Sciencewise Theory of Change for Strategic Planning 2014-2015* (Sciencewise, 2014) and *Evaluation of Sciencewise-ERC* (Sciencewise, 2011).

- **EU Regulatory Fitness and performance Programme (REFIT) (European Commission, 2015).** The implementation of REFIT involves comprehensive policy evaluations aimed at assessing whether the regulatory framework for a particular policy sector is 'fit for purpose'. The so called 'Fitness checks' provide an evidence-based critical analysis of whether Union actions are proportionate to their objectives and delivering as expected. Recent relevant evaluations under REFIT include the following, while a number of evaluations are already planned for the future as set out in the *Commission's Forward Planning of Evaluations and Studies: 2016 and beyond*[6]:
  1. **Food:** Fitness Check on the General Food Law Regulation (2015)
  2. **Energy:** Evaluation of Renewable energy legislation (2015)
  3. **Environment:** Fitness Check of EU Waste Legislation (2014), Evaluation of the Environmental Noise Directive (2016 ongoing)
  4. **Climate Action:** Evaluation of the Carbon Capture and Storage(2014)

- **Better Regulation Guidelines on Impact Assessment (European Commission, 2015).** An Impact assessment is envisaged to happen as early in the process of policy development as possible. As such, it can provide valuable input to evaluations offering descriptions of the problem, objectives, policy options, related issues and relevant impacts, providing an understanding of ex-ante uncertainties, possibilities and information sets (also see Jaffe et al., 2005).

- **The Rainbow Framework (BetterEvaluation, nd).** This framework organises evaluation methods into seven clusters of evaluation tasks and describes the questions that need to be answered under each task. It acts as a checklist to ensure all factors have been considered in the evaluation design, implementation and dissemination, while the website offers a useful resource for evaluation practitioners on the different methods, approaches, data collection, analysis and reporting relevant to evaluations.

[6] http://ec.europa.eu/info/files/plans-evaluations-and-studies-2015-and-beyond_en

.

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