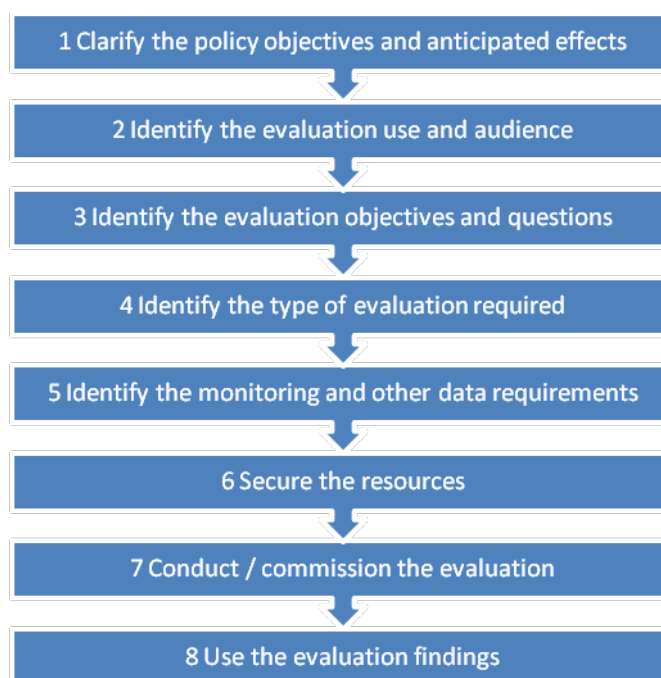# DECC Evaluation Guide

Aim of this guide

This Guide has been developed to help anyone who is responsible for planning, commissioning or managing an evaluation of a policy, programme or project. It sets out the main steps involved in planning and undertaking an evaluation; provides an **Evaluation Plan Template** which relates to the steps; and offers key questions to be considered at each step.

The DECC **Policy Evaluation Team** are available to support and advise policy teams planning and undertaking evaluations. The HMT **Magenta Book** is the cross-government guide to evaluation and provides further information and technical advice (see Further Information section of this guide).

The steps involved in planning and undertaking an evaluation are:

1 Clarify the policy objectives and anticipated effects

2 Identify the evaluation use and audience

3 Identify the evaluation objectives and questions

4 Identify the type of evaluation required

5 Identify the monitoring and other data requirements

6 Secure the resources

7 Conduct / commission the evaluation

8 Use the evaluation findings

Articulating each of these steps in an **Evaluation Plan** will help identify the intended impacts of a policy, and set out how you will assess whether and how these have been achieved. Evaluation planning should be undertaken early, as part of policy/programme **Business Case** and **Delivery Plan** development. This is because the design of a policy will affect how rigorously it can be evaluated, and it will almost always be necessary to collect some data for the evaluation **before the policy is implemented** and throughout its delivery.

A summary of each step and link to the relevant section of the guide is presented overleaf.

Step 1: Clarify the policy objectives and anticipated effects - page 6

Identifying the objectives and anticipated effects of the policy will provide a framework for the evaluation plan, and help identify exactly what the evaluation should be assessing

Step 2: Identify the evaluation use and audience - page 9

To ensure the evaluation provides useful evidence, it is important to consider the audience and their requirements *before* it is started, as this will influence its design

Step 3: Identify the evaluation objectives and questions - page 11

Building on the first two steps, it is essential that the questions the evaluation needs to address are set out as specifically as possible, relating to what the policy has achieved, how it was delivered and how it contributes to wider objectives

Step 4: Identify the type of evaluation required - page 13

Reviewing the evaluation questions will help inform the type of evaluation required; impact (what difference did the policy make) or process (how was it delivered), or both

Step 5: Identify the monitoring and other data requirements page 17

Data requirements should be identified early and collection processes built in from the start, to ensure baseline and monitoring data are collected before and during policy delivery

Step 6: Secure the resources - page 21

Judgement needs to be made about the scale and form of the evaluation, including whether it should be commissioned externally, to inform resource allocation and business planning

Step 7: Conduct / commission the evaluation - page 24

A dedicated DECC project manager should be responsible for commissioning (if externally commissioned) and managing the evaluation, following DECC standard processes

Step 8: Use the evaluation findings - page 26

Evaluations are designed to provide useful information to inform current and future policy development, so early planning for dissemination is essential
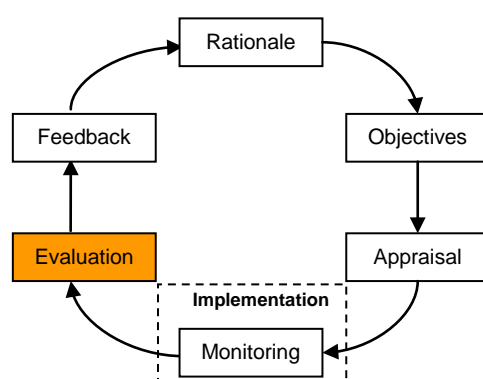
Introduction

What is evaluation?

Evaluation is the robust analysis of how a policy has been delivered <u>in practice</u> and whether its objectives were met. Evaluation can provide an understanding of the *actual* economic, financial, social and environmental impacts of a policy or programme, and can provide an assessment of how the policy was implemented, why it did (or did not), deliver as expected, and whether it represents value for money. Good evaluation is multi-faceted, looking at all aspects of the policy, and should look for any evidence that a policy does not work, as well evidence to show that it does.

This information can inform ongoing implementation decisions to maximise the impact of policies and can be fed into future policy development decisions. It is an essential component of the Policy Cycle, as shown in figure 1.

Figure 1 The ROAMEF policy cycle.

Source: Green Book (modified)

Evaluation should not be confused with:

- <u>Appraisal and modelling</u>: Appraisal and modelling examine *expected* costs, benefits and other impacts. Evaluation, in contrast, looks at what *actually* happened in practice.

- <u>Monitoring and performance management</u>: Monitoring data collected as part of ongoing policy/programme delivery (to assess performance, keep track of the operational delivery and assess whether it is running effectively) is a key component of any evaluation. However, an evaluation typically goes beyond this to look at measuring the impact of a policy compared with what would have happened without it, and to develop an understanding of why.

Why evaluate?

For most policies, it is essential to know the impact they have had in practice (in contrast to the expected impacts) and understand what worked or didn't work. Such evidence can help DECC alter elements of the policy to maximise impact and inform future policy development in the area. With ambitious targets stretching 40 years ahead, it is essential that we have a solid evidence base of <u>what works, by how much, and why,</u> to build upon. In addition, in times of fiscal constraint, it is all the more important to be able to understand and demonstrate value for money, and to inform decisions on the best use of scarce resources.

DECC is also subject to external scrutiny – the NAO and Public Accounts Committee scrutinise what our policies and programmes deliver, and how we demonstrate this. The Committee on Climate Change, HM Treasury, lobbying bodies, stakeholders and the public will all hold us to account and ask us to demonstrate that our policies and programmes are effective, and that we have robust systems in place for evaluating effectiveness.

What should be evaluated?

Evaluation needs to be proportionate to the risks, scale and profile of the policy. <u>High risk or high uncertainty</u> innovative policies are likely to require a reasonably comprehensive evaluation to understand how they are working in practice, and whether they are delivering the predicted impacts. Any policy or programme <u>set up as a pilot or demonstration</u> should be fully evaluated. Equally, <u>large scale or high-profile</u> policies are likely to require robust evaluations to build the evidence base about what works, and for accountability purposes: to demonstrate that public money has been well spent. All Impact Assessments now require Post Implementation Reviews (PIRs), 3-5 years after implementation: essentially an evaluation. <u>Small or routine</u> policies, for which evaluation is unlikely to generate many learning benefits, may only need light touch monitoring/ evaluation to confirm they have been delivered as intended.

The planning phase of *all* policies and programmes should include consideration of what type and depth of evaluation is appropriate. Although a comprehensive evaluation may not always be called for, it will **always** be important to assess some aspects of the delivery process. This will, at the very least, include some degree of **monitoring**, to track the inputs and outputs that are actually delivered. However, simply counting the inputs and outputs risks coming to faulty conclusions (observed changes might have occurred anyway, without the policy intervention) or missing the valuable evidence that can be gained from evaluation: understanding the overall impact of the programme, how it worked, for whom and why, and whether there were any unpredicted benefits or problems. Such evidence is increasingly important in demonstrating accountability, informing future policy development and helping inform future allocation of scarce resources.

When to think about evaluation and monitoring

Good evaluation is built into the design of a policy or programme and thought about throughout its development and implementation. This allows both the policy and the evaluation to be tailored to maximise the potential for robust, usable findings that can help future decision making. **The design of a policy will affect how rigorously it can be evaluated.** Failure to think about evaluation and build this into the policy design can preclude a reliable understanding of the impact of the policy being achieved: appropriate baseline data is often not collected and a comparison/control group is not available to help understand what would happen in the absence of the policy.

This is particularly important for pilot/demonstration programmes which, by their very nature, are designed to generate learning and understanding about the potential for scaling up or replication in other areas.

There is a requirement in DECC to consider evaluation as part of programme **Delivery Plans and Business Cases**, and those scrutinising these documents (e.g. the Approvals Committee) will expect to see a full consideration of what the evaluation needs are, how they will be built into the policy delivery, and the resources (programme and admin) that will be required to deliver a successful evaluation. Using the Evaluation Plan template will help to identify what is needed, when and why.

Designing and conducting an evaluation *after* a policy has been implemented does happen, but there are risks (a robust evaluation might not be possible) and the quality of the evidence is likely to be poorer. It will be necessary in this situation to identify what relevant information is available, e.g. data collected as part of ongoing performance monitoring and/or administrative data. **However, this often misses baseline data, collected before the policy was implemented, or data collected for a comparison or control group**. This will result in a less robust assessment of the policy being possible, and key questions, such as 'would the outcomes have happened anyway?' being unanswerable. Nonetheless, it will normally be possible to undertake some assessment of the delivery process and immediate outputs of a policy if the evaluation is planned and undertaken after policy implementation.

Developing an Evaluation Plan will undoubtedly be an iterative process: it is likely to be necessary to re-review the evaluation objectives and questions once the resources available for the evaluation and the intended use/audience of the results have been clarified. The DECC Policy Evaluation Team are on hand to provide advice.

Developing an evaluation plan

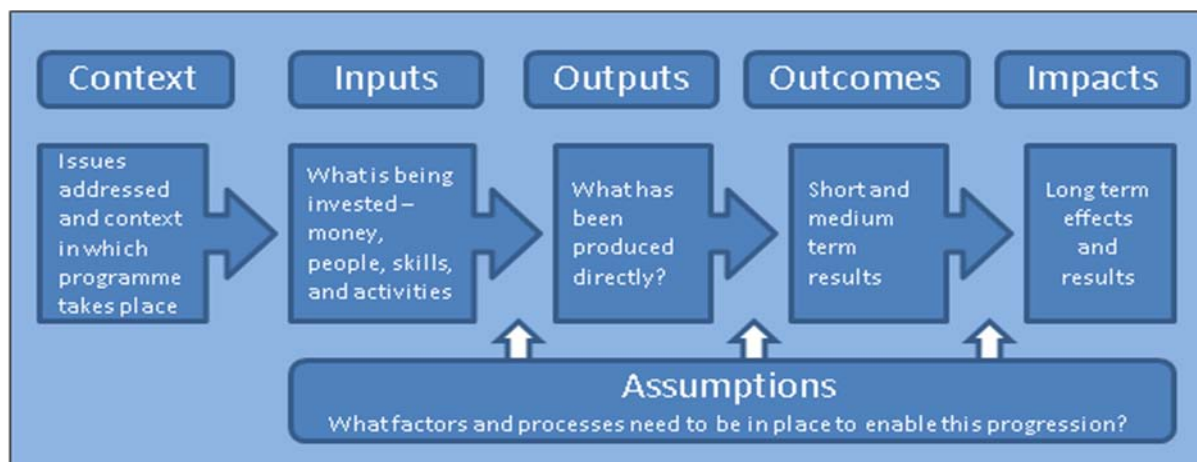## 1. Clarify the policy objectives and anticipated effects

The first thing to do when planning an evaluation is to identify the objectives and anticipated outcomes and impacts of the policy or programme. It is important to set this out clearly, to provide a common understanding and framework for the evaluation plan, and help identify exactly what the evaluation is assessing.

The recommended way to do this is to set out the logic and assumptions of the programme or policy (often called the 'logic model', 'logic map' or 'programme logic'), as shown in figure 2. For complex policies, it may be necessary to develop a series of linked logic models. A similar exercise may have been undertaken as part of the Business Case development and can be used as a starting point.

The logic model identifies the following elements:

- The **issues** being addressed and the **context** within which the policy takes place.
- The **inputs**, i.e. the **resources** provided and **activities** undertaken by DECC and/or delivery bodies.
- Initial outputs of the policy (e.g. target groups engaged, grants delivered, information provided).
- The **outcomes** of the programme (i.e. short and medium-term results, such as uptake of energy efficiency measures, knowledge generated, components manufactured).
- The anticipated **impacts** (e.g. long-term results such as behaviour change, jobs created, energy and greenhouse gas emission reductions).
- The **assumptions** made about how these elements link together and enable the programme to successfully progress from one element to the next (e.g. consumers take up offers; firms are aware of and take up loans; new technologies are used etc).

Figure 2: Logic model of the programme theory

6

An evaluation will seek to identify the extent to which each of these elements did or did not materialise in practice, and will usually also **examine the links between each element**, to understand the progression between them and the delivery process.

The logic model can also be used to help develop success criteria or key performance indicators for a policy, by establishing specific targets related to the outputs and/or outcomes – for example X energy efficiency measures installed by a certain date.

Logic models are read from left to right, leading you through a time sequence from the initial concept, through implementation to short term and longer term effects and results. However, they are generally developed from right to left. In other words, it is best to start with the impact or change that the policy is intended to achieve (i.e. policy objectives), before working backwards through the steps required in order to achieve these objectives.

Producing the logic model can be undertaken as a desk exercise, based on a review of policy or programme documentation. However, developing and testing the model with key stakeholders – for example organisations involved in the delivery and representatives of different groups affected by the policy, will help ensure all the important issues are identified and assumptions are reality checked.
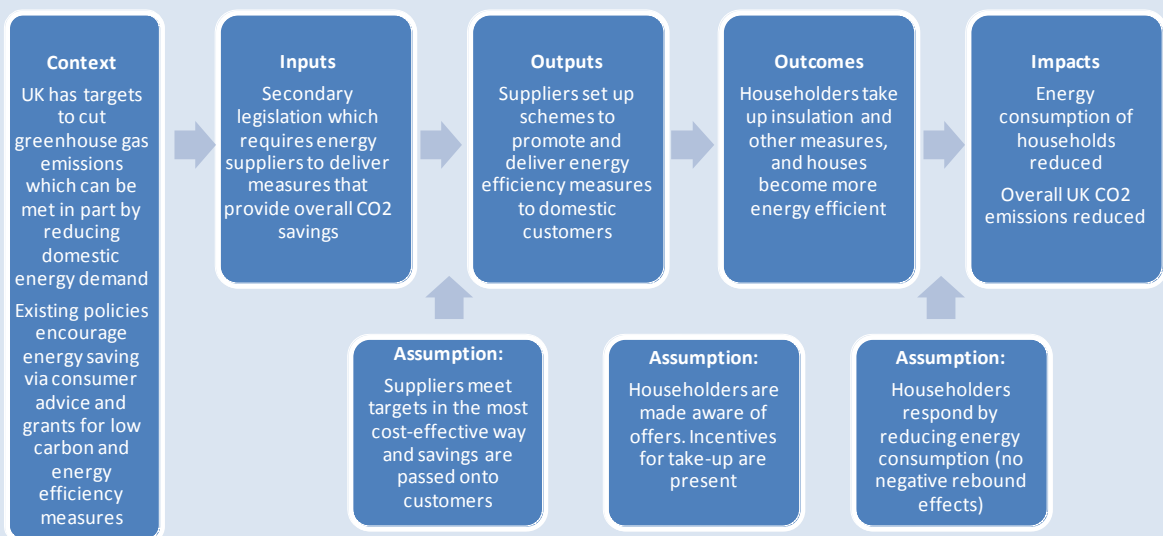

EXAMPLE STEP 1: Clarifying the policy objectives and intended effects

The following example illustrates how step 1 could be applied to a hypothetical DECC policy to reduce domestic energy consumption via an obligation on energy suppliers.

The **objectives of the policy** set out in the business case / impact assessment would be used as a starting point for the evaluation – in this case to improve the energy efficiency of homes and contribute to UK greenhouse gas emissions reductions.

Key **contextual factors** would then be clarified, including who the policy is intended to reach, and other related and/or overlapping policies already in place. In this case, previous schemes to promote domestic energy efficiency and current policies to incentivise uptake of small-scale renewable technologies.

This information would then enable an initial **policy logic model** to be developed, before being refined and verified with key stakeholders. A simplified logic model for this policy would look something like this:

Context	Inputs	Outputs	Outcomes	Impacts
UK has targets to cut greenhouse gas emissions which can be met in part by reducing domestic energy demand	Secondary legislation which requires energy suppliers to deliver measures that provide overall CO_2 savings	Suppliers set up schemes to promote and deliver energy efficiency measures to domestic customers	Householders take up insulation and other measures, and houses become more energy efficient	Energy consumption of households reduced
Existing policies encourage energy saving via consumer advice and grants for low carbon and energy efficiency measures				Overall UK CO_2 emissions reduced

Assumption: Suppliers meet targets in the most cost-effective way and savings are passed onto customers	**Assumption:** Householders are made aware of offers. Incentives for take-up are present	**Assumption:** Householders respond by reducing energy consumption (no negative rebound effects)

The questions in table 1 will help with the development of the policy logic model.

Table 1: Questions to help develop the logic model

context and issues	• What are the stated objectives of the policy? • What contextual factors may influence the ability of the policy to achieve its outcomes and impacts? • Who will the policy affect, both directly and indirectly? • What is the existing knowledge base? • What other policies are in place and may also affect the expected outcomes?
impacts	• What is the overall goal of the policy? • What is the policy aiming to achieve in the **long term**? (e.g. behaviour change, widespread use of new technology; energy savings; GHG emission reductions) • What policy objectives will it address?
outcomes	• What is the policy expected to achieve in the **short to medium term**? • What kind of changes would you expect to see as a result of activities? (e.g. take up of energy efficiency measures; skills increases; new low carbon technologies developed)
outputs	• What will be **delivered as a direct result** of the policy? (e.g. grants provided or leveraged; energy efficiency /low carbon measures installed) • What activities will directly result from the policy? • What participation will directly result from the policy?
inputs	• What **financial resources** are being invested by DECC and/or other partners? • What **activities** will be undertaken to deliver the policy by DECC and delivery partners? • What **other resources** are being invested? (e.g. people and organisations, skills, research, etc.)
Assumptions	• What is assumed about how inputs will feed into outputs? (e.g. consumers are engaged / informed; loans are taken up; investors can be identified) • What is assumed about how outputs will feed into outcomes? (e.g. training exists to increase skills; new product components can be sourced; energy efficiency measures are taken up) • What is assumed about how outcomes feed into impacts? (e.g. new technology is used; rebound effects are less than CO_2 savings)

2. Identify the evaluation use and audience

You should have a clear idea from the start about the intended use and audience for the evaluation: what will the findings feed into and inform? when is the information required by different stakeholders? This will help direct all stages of the evaluation and ensure maximum impact from the findings and should be set out in the Evaluation Plan.

To ensure the evaluation provides useful evidence, it is important to consider the requirements of the users of the findings *before* the evaluation is started so the evaluation can be designed to answer these specific question and deliver findings when they are needed. For example, the findings may be needed to improve the ongoing delivery process, inform specific decision-making, clarify thinking, provide accountability to stakeholders and contribute to improved knowledge amongst those best able to take advantage of it.

Thus, when developing the evaluation plan, it is important to understand:

- **Who** the target end-users of the findings will be - e.g. programme managers and the programme board; other policy makers and analysts within DECC; other Government Departments; delivery bodies; key stakeholders including industry bodies and energy suppliers, local community groups etc.
- **When** findings are required, and the timing of decisions they need to feed into.
- **What form of evidence** is required - e.g. a quantitative cost-benefit assessment of impacts may be required by HM Treasury, while detailed information about effective delivery mechanisms may be sought by programme managers of similar programmes.
- **What form of communication** would be most appropriate to ensure findings are useful and used, and reach the target audience and can provide maximum impact.

These considerations will influence the design of the evaluation. By understanding the range of requirements for the evaluation and their relative priority, the evaluation can be tailored to generate the relevant evidence to the required timescales, and/or (just as importantly) a decision can be taken early about the questions which can realistically be answered in the desired timescales.

The questions in table 2 will help you to think about how the findings will be used, and by whom.

Table 2: Questions to help identify how the evaluation findings will be used

- Who will be the main users of the findings?
- When are the findings required?
- What will the findings be used for?
- How will you ensure the evaluation findings make an impact?
- How will the findings be shared and disseminated?
- Who else should be made aware of the findings?

EXAMPLE STEP 2: Identifying the evaluation use and audience

Continuing the example of a policy to reduce domestic energy consumption, consideration of step 2 would look like this:

o The key **users of the findings** would be the DECC policy team, colleagues leading on related policies, relevant other government departments (Defra, CLG and devolved administrations) and the regulator for the scheme (Ofgem). Other key stakeholders, including energy suppliers and partners involved in delivering the scheme, will also have an interest.

o The **findings will be used to** (a) assess whether the policy is delivering as expected and to inform any changes to the policy and/or the delivery model over time, to ensure it is delivered as effectively as possible and (b) to understand the overall impact of the policy for accountability purposes and to inform future energy supplier obligations.

o The **timing of the findings** needs to fit with the agreed review milestones as laid down in the Act that created the policy, and with the proposed policy development of a new energy supplier obligation and a new Energy Act.

o The **findings will have maximum impact** by having key policy leads on the steering group for the evaluation; the evaluation project manager providing the steering group with regular updates of progress and emerging findings; multiple channels of communication being used for the final findings, including a published final report, seminars, workshops with industry groups and other government departments, one-page summaries for other key stakeholders and press releases (Press Office will be kept informed of progress).

3. Identify the evaluation objectives and questions

The third step is to clarify what questions the evaluation needs to address.

All policies will have specific questions that the evaluation will need to answer, about what the policy has achieved, and how it was delivered. It is important to think about these *before* more detailed consideration of data requirements: you want the questions to shape the data collection, not the other way round. Returning to the logic map produced at step 1 should help you to formulate specific evaluation questions, since this will have identified the anticipated outcomes and impacts and the underlying assumptions that might need tested.

There will be high level objectives common to many policies (e.g. DECC priorities), and evaluations should generally seek to demonstrate the extent to which a policy contributes to these objectives. It may also be necessary for an evaluation to demonstrate the extent to which a policy contributes to carbon budgets and 2050 targets. This is discussed further in section 5 of this guide.

The questions in table 3 will help with development of the evaluation questions.

Table 3: Issues to think about when developing evaluation questions

What difference did the policy make? (outcomes and impacts)	• How will you know if the policy is a success? • Who is expected to benefit from the policy? • Which of the anticipated impacts and outcomes are most important to assess? • What longer term impacts should be assessed? • Do you need to quantify impacts, as well as describe them? • Do you need to assess the actual cost-benefit of the policy? • Are you expecting the policy to deliver wider knock-on effects, and is it important to assess these?
What was delivered? (inputs and outputs)	• How will you know if the policy is on track? • Are there interim steps in delivery which can be identified? • What is being delivered by external organisations / supply chains?
How was the policy delivered? (processes)	• What do you need to learn about the delivery process? • Do you need to understand why the policy does/does not achieve anticipated outcomes? • Which aspects of the delivery process are innovative or untested? • Is it important to learn about uptake (e.g. of energy efficiency measures or new technology) and customer attitudes ? • What contextual factors might affect delivery (e.g. economic climate, other policy measures, innovation developments, etc)?

Identifying the evaluation questions will likely be an iterative process. It may be helpful to start with an 'ideal' list of questions for the evaluation, but you will need to consider what is proportionate to address, the resources available and ultimately what will be possible or not. As you develop the evaluation questions, think about *why* each question is important, and how the information will be used.

<div style="border: 2px solid; padding: 10px; background-color: #dce6f1;">

EXAMPLE STEP 3: Identifying the evaluation objectives and questions

The key **evaluation questions** for the example policy are identified by policy and analyst leads within DECC and the key stakeholders (Ofgem). The questions are also verified with wider stakeholders.

Questions related to **what difference did the policy** make include the following:
- What is the impact of the policy on take-up of energy efficiency measures and change to energy use behaviour?
- How does this differ amongst different groups of people and different types of housing?
- What are the actual costs and benefits in terms of greenhouse gas emissions reduction?
- What are the wider economic impacts on innovation, jobs etc?

Questions related to **what was delivered** include the following:
- How many schemes are set up to deliver energy efficiency measures, and where are they based?
- How many different energy efficiency measures are taken up by customers, when and where?

Questions related to **how was the policy delivered** include the following:
- What are the different delivery models that have been adopted in different areas, and what are their relative strengths and weaknesses?
- What factors prompt take-up of energy efficiency measures by customers?
- How is delivery monitored and quality assured?

</div>

4. Identify the type of evaluation required

Very broadly speaking, there are two different types of evaluation, addressing different types of evaluation questions:

- **Impact evaluation** assesses the outcomes and impacts generated as a result of the policy. This type of evaluation addresses the question 'what difference did the policy make?'
- **Process evaluation** examines how, why, and under what conditions a policy and its outcomes were delivered. This type of evaluation addresses the question 'how was the policy delivered?'

In practice, an evaluation will often include both impact and process elements. Both are required to understand what difference a policy did (or did not) make, why, and how. Monitoring what was delivered directly by the policy should *always* be undertaken, and will feed into both process and impact evaluation.

What difference did the policy make? Impact evaluation

A key objective of an **impact evaluation** is to assess whether any changes occurred after the policy was implemented (the observed outcomes) and whether these were a result of the policy or would have happened anyway. Impact evaluations are typically undertaken to account for use of resources, to demonstrate and quantify the benefits achieved by a policy, and to aid decision making about future resource allocation.

To make this assessment it is necessary to understand what would have happened in the absence of the policy – this is known as the **'counterfactual'**. Establishing the counterfactual is not easy, since by definition it cannot be observed – it is what would have happened if the policy had not gone ahead. For this reason it has to be estimated or constructed. The strength of an evaluation design and the quality and availability of data will determine the extent to which the counterfactual can be robustly established.

The most robust counterfactual is typically achieved through random allocation, where it is by chance whether an individual or group is assigned to either the 'treatment' (i.e. receives or is affected by the policy) or counterfactual group. However, this is not always possible to undertake in a policy evaluation, in which case other methods can often be employed. A weak or non-existent counterfactual will result in there being uncertainty as to whether the outcomes would have happened anyway, regardless of the policy, and it will therefore not be possible to say conclusively whether the policy was effective or not or to quantify its impact.

Another way of looking at this is that any observed outcome will likely be affected by a number of factors in addition to the policy (including other DECC policy interventions). It is important for an impact evaluation to demonstrate, as far as possible, that the effects were

caused by the policy or programme. This is known as **attribution**. If evaluations are not designed to show that the policy has caused or influenced a change, then uncertainty will remain as to whether the observed effects were actually caused by other factors.

It will often be important for an evaluation to assess whether there were any unintended effects of a policy, either positive or negative (for example, a policy to reduce household energy consumption through improving energy efficiency might result in a net increase in carbon emissions if people respond by having warmer homes, or spending the money saved from reduced heating costs on additional electrical appliances, air travel etc). A policy may also have **wider effects** which are not so easily observed, such as knock-on effects in the local economy (known as multiplier effects). Assessment of such effects can be built into an impact evaluation although they would generally only be assessed in the evaluation if they are anticipated in the business case or appraisal.

An impact evaluation will generally seek to describe and quantify the outcomes and impacts of the policy, as far as possible. It may also be appropriate to monetise these impacts, which can then be compared with actual costs, as part of an economic evaluation or **cost-benefit analysis.** This assessment should be proportionate to the scale and profile of the policy, and will usually only be undertaken as part of a comprehensive impact and process evaluation. Cost-benefit analysis is described further in the box below.

Impact evaluation and cost-benefit analysis

Cost-benefit analysis seeks to quantify in monetary terms as many of the costs and benefits of a policy or programme as feasible. The HM Treasury Green Book provides guidance for imputing monetary values to inputs, outputs and outcomes. DECC guidance on valuation of energy usage and greenhouse gas emissions should also be consulted (see further information section for links).

Cost-benefit analysis is undertaken both as part of an **appraisal** – to support a government decision to adopt a new policy or programme which would result in benefits and/ or costs to the public; and as part of an **evaluation** – to retrospectively analyse the costs and benefits which actually materialised as a result of the programme. Generally, monetisation assumptions used as part of the appraisal can also be applied in the evaluation.

Costs and benefits should normally be based on market prices as they usually reflect the best alternative uses that the goods or services could be put to (the opportunity cost). Where there are market failures some adjustment may be made to market prices, or alternative valuation methods used if there is no market price – such as for carbon abatement in the non-traded sector.

Costs and benefits which are not monetisable also need to be brought into any assessment (e.g. qualitative description of the benefit). In such cases it is important that any assessment does not just rely on the net present value (NPV) figure (monetised benefits to society minus monetised costs to society), but also considers the non-monetised costs and benefits.

How was the policy delivered? Process evaluation

Process evaluations seek to assess whether a policy is being implemented as intended, and what is working well and less well in the delivery. Process evaluations are important for determining the key success factors (and barriers) for effective delivery of a policy, and provide information about how particular aspects of the policy contribute to the observed outcomes. Process evaluations are undertaken to help improve delivery during implementation, to generate learning about how to deliver future policies, and/or to aid interpretation of impact evaluation findings (see below).

Process evaluation involves examining the different steps involved in the design and delivery of the policy and how they link together – e.g. stakeholders and supply chains involved; partnership arrangements; decision-making processes etc. A process evaluation usually includes qualitative research with different stakeholders (e.g. focus groups and in-depth interviews). It may also involve examination of organisational data, key documents and events. Information from different sources, people and time-points is usually 'triangulated', to obtain a balanced and well-informed understanding of the situation. For example, assessing the policy at different stages during and after roll-out; using information from different sources; and using more than one approach or method to answer a specific question.

Using impact and process evaluation together

An evaluation will often include both impact and process questions. Determining whether or not a policy has had an impact often involves asking questions about how it has done so, for whom, why, and under what conditions it had the effect. When combined with an impact evaluation, a process evaluation can help explain why the observed outcomes of a policy or programme were different to that expected. Without this there is a risk of not knowing whether a policy failed to deliver the anticipated impact because of a failure of the policy design, or failures in the way it was implemented.

It may also be important for an evaluation to assess any unintended effects of a programme, either positive or negative. Again, a process evaluation can help inform this, for example by identifying who an initiative actually reached. However, as noted in the previous step, it is important to think about what questions can practically and feasibly be included in the evaluation as it will not always be possible to undertake a full impact and process evaluation.
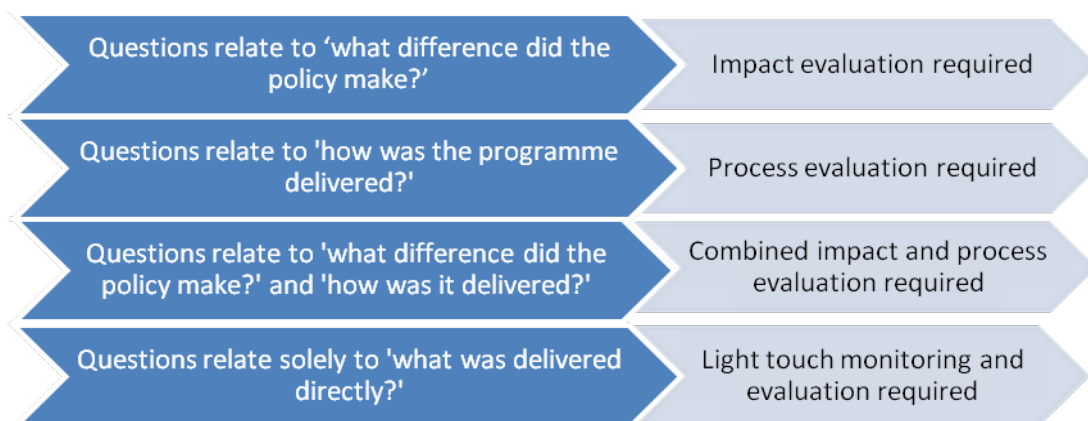
Deciding on the type of evaluation

The initial step should always be to clearly articulate the objectives and questions for the evaluation; decisions on the evaluation methodology then follow. Following the steps set out in this guidance in order will help with this.

Having a clear idea about the required type of evaluation at an early stage will also help inform the design of the policy to ensure it is 'evaluable' (e.g. if an impact evaluation is required, how will the counterfactual be assessed?), the scope of the evaluation, and the expertise required. Such decisions should involve both policy and analytical colleagues and be made **well before** thinking about commissioning an external evaluation.

Reviewing the evaluation questions identified at step 3 will help inform the type of evaluation required, as shown in table 4. The scale and depth of the questions will inform whether the evaluation needs to be a full impact and process evaluation, a light touch assessment, or somewhere in-between. As stated earlier, large-scale, costly or innovative policies will be more likely to require a more comprehensive evaluation. For policies that have already been delivered, and do not have high-quality data associated with them, it is unlikely that an impact evaluation can be undertaken.

Additionally, the level of activity (i.e. project, programme or policy) and the availability of data and/or resources may constrain the number of questions an evaluation can realistically answer and type of evaluation which can be conducted. This is discussed further at step 6.

Table 4: Identifying the type of evaluation

Questions relate to 'what difference did the policy make?'	Impact evaluation required
Questions relate to 'how was the programme delivered?'	Process evaluation required
Questions relate to 'what difference did the policy make?' and 'how was it delivered?'	Combined impact and process evaluation required
Questions relate solely to 'what was delivered directly?'	Light touch monitoring and evaluation required

---

EXAMPLE STEP 4: Identifying the type of evaluation required

The evaluation questions identified at step 3 indicated evidence needs for the example policy, in terms of understanding **what difference the policy made**, **what was delivered** and **how the policy was delivered.** In addition, it is a high-profile and innovative policy, and one that will inform future policy decisions in this area. The policy is in development, meaning there is an opportunity to shape the design of the policy to obtain high-quality base line information, and data from a counterfactual group. It is however not possible to randomly assign people to 'treatment' or 'counterfactual' groups. It will, however, be important to test different delivery models in early roll-out to inform later delivery.

In this case, a combined process and impact evaluation is required, and can be justified given the high-profile nature of the policy and the use that will be made of the resulting evidence. Constructing a counterfactual will be challenging, but building consideration of evaluation into policy design will mean it is a challenge that can be met, in this case by using phased roll-out across the country, so similar groups both get the policy, but at slightly different times, so that the groups can be compared.

5. Identify the data requirements

The evaluation questions and type identified at stage 4 will inform the data requirements: impact evaluations will require different types of data to process evaluations. In every case, the evaluation data requirements should be considered alongside all the other data requirements of the policy (e.g. monitoring and performance data) and existing data should be utilised and built on where possible.

Monitoring and performance data

Monitoring and evaluation are complementary activities, and ideally the design and requirements for each should be considered together, so the comprehensive data needs of the policy can be considered in the round. This will facilitate the collection of relevant and high quality data and avoid duplication or missed opportunities for the collection of key data. Early identification of any existing data or other ongoing data collection processes that can be utilised for the evaluation will ensure best use of resources and effort.

Monitoring and performance data relate to information collected and used as part of ongoing policy delivery. Such data are used to help the operational delivery and performance of a policy or programme, and assess whether it is running effectively, e.g. by checking progress against interim targets. Collection of monitoring data should be designed to meet the needs of both the policy delivery and the evaluation – data which is directly useful to those collecting it tends to be better quality than data collected solely for the purpose of research. *All* policies should have some degree of monitoring undertaken.

Evaluation data

A good evaluation relies on good data. In addition to monitoring data, this would often include new data collected specifically for the evaluation, and/or data collected for other purposes (secondary data). It is important to note that an impact evaluation will require data to assess the counterfactual, i.e. data from a comparison or control group (e.g. firms or households) who are not affected or in receipt of the policy intervention. Information from such groups who are receiving no direct benefits can often be challenging to obtain.

The data required for an evaluation will be related to assessing the inputs, outputs, outcomes and impacts of the programme. It will also be related to testing how these elements are linked together (i.e. the processes and assumptions). Generally, baseline (i.e. pre-policy) data will be required in order to show what changes have occurred since the policy has been implemented, for both 'treated' and comparison / control groups.

This is further explained, as follows:

- **Monitoring and performance data** used in evaluation includes information on delivery of inputs and outputs – e.g. loans offered and taken up; investment leveraged, energy efficiency measures provided and installed etc. This will usually be quantitative data, but some aspects of the programme may be monitored qualitatively.
- **Quantitative evaluation data** would normally be collected in order to assess change in outcomes in both the 'treated' and counterfactual groups – e.g. surveys conducted before and after policy implementation to assess change in attitudes and behaviour. This will usually be collected as part of a specific policy evaluation, but in some cases may be collected and collated through wider analytical work such as the National Energy Efficiency Data framework (see below).
- **Qualitative evaluation data** may be collected to assess whether anticipated outcomes and impacts have materialised (those which cannot be assessed quantitatively); identify any unintended outcomes; and examine the processes involved in transforming inputs into outputs and outcomes – e.g. interviews, workshops and observations.
- **Secondary data** can be used to provide information about the wider context in which the policy is being delivered – e.g. national data on GDP, public attitudes, fuel use, prices etc, to help understand how external factors have affected outcomes.

Timing of data collection

Monitoring and performance data requirements should be built in from the start of any policy, so that collection processes are established alongside policy delivery. This may involve building the collection of monitoring and performance data, to required standards, into delivery arrangements with external partners. Ideally, additional data requirements for the evaluation will also be identified early, before policy implementation commences. Relevant data would then normally be collected before the policy starts (baseline), during the delivery (interim), and after completion, to allow before/after change to be assessed. The exact length of time for collection of 'after' data should be aligned with when outcomes and impacts are expected to materialise.

It may be possible to address some evaluation questions using post-hoc data collection, but this tends to result in weaker designs and less robust findings, for example asking people retrospectively to remember what they did or thought sometime before the policy was delivered is an unreliable method of measuring behaviour or attitude change.

The cost of undertaking an evaluation will also be substantially lower if it can use relevant data produced through monitoring activities, rather than necessitating extensive new data collection.

Identifying data

You should first consider what sources of data already exist and may be appropriate for the evaluation. You may need to weigh up the quality and usefulness of existing data (e.g. Local Authority or administrative data) against the practicality and feasibility of collecting new data to assess outcomes.

Advice should be sought from **DECC statisticians** at this point.

If new data is required, you will need to decide how this should be collected, and by whom. For example, will it be by those delivering the policy, or as part of a monitoring / evaluation contract? This also needs to be built into the resource considerations (see step 6).

One important source of robust data for evaluating the impact of national energy efficiency policies is DECC's national data set of domestic properties in England and Wales – the **National Energy Efficiency Data framework** (**NEED**). This combines at property level, data on annual electricity and gas consumption (covering all 26 million homes since 2004), with data on housing type, installed energy efficiency measures, and socio-demographic indicators. Analysis of NEED is undertaken by DECC analysts, and will be likely to form a key part of the evaluations of key DECC energy efficiency policies.

Core evaluation measures

Although the exact monitoring and data requirements will be specific to an evaluation, there will be high level DECC objectives common to many policies and programmes – e.g. reducing energy consumption and greenhouse emissions. As far as possible, consistent measures should be used in evaluations to assess the extent to which common outcomes have been achieved across different policies, programmes or projects. This includes key performance indicators (outputs), as well as measures relating to policy outcomes and impacts. This ensures any evaluation seeking to assess the contribution a policy made to high level objectives (e.g. DECC priorities) does so in a consistent way, and allows assessment of the relative contribution of the policy.

Advice should be sought from the **Policy Evaluation Team** when identifying the measures and data requirements for any evaluation.

It is also important that any assumptions – for example for carbon and energy prices or emission factors – are applied consistently across different evaluations and between the appraisal or business case and the evaluation. The CO_2 equivalent change in greenhouse gas emissions resulting from the policy or programme should always be recorded in million tonnes of CO_2 equivalent.

Advice should be sought from **DECC economists** and the latest **DECC guidance on valuing emissions** should be used (see **Further Information** section).

The questions in table 5 should help identify what data is required for the evaluation.

Table 5: Questions to help identify data requirements for the evaluation

- What data is required to monitor relevant outputs and outcomes?
- What data is already collected or available, and how reliable is it?
- What additional data needs to be collected?
- Who should be responsible for data collection?
- What processes need to be set up to collect monitoring data?

EXAMPLE STEP 5: Identifying the data requirements

For the energy company obligation policy, the following data sources will be required:

- The **monitoring data** will need to include information to feed into the evaluation about when and where measures are delivered, and taken up, by individual households.
- A **baseline population survey** will be required to give an indication of the starting position for energy use behaviour, and repeated after policy implementation to assess change amongst households who have, and have not, taken up measures.
- **Qualitative research** with stakeholders will be required at different points in time during the policy delivery, to provide an understanding of the supply chain from different perspectives – which aspects and delivery models are working well or less well; what are the key drivers and barriers. Qualitative research with customers will be required to provide understanding about why they have taken up measures delivered through the scheme.
- **Quantitative data** relating to change in actual annual household energy consumption is also required – this is **already being collected** and held by DECC via NEED.

6. Secure the resources

As already mentioned, judgement needs to be made about the scale and form of evaluation that is required for a particular policy, including whether it should be commissioned externally or conducted (either partly or wholly) in-house. In some circumstances, it may be useful to undertake a scoping or feasibility study to support this decision-making process. This form of assessment can foster greater understanding of what can and cannot be evaluated, and therefore what level of investment is required, and can support the development of an appropriate evaluation design for large or complex evaluations.

All evaluations, even those commissioned to an external contractor, will require significant internal input to ensure they are designed and delivered successfully. A number of resources will need to be considered, as follows:

Financial resources: It is important to think about the financial resources required for evaluation early, as part of budget planning and business case for the policy or programme. In particular, it should be noted that a substantial part of the evaluation costs can often be incurred *after* the programme delivery has completed. It is not possible to give a fixed sum or proportion of budget for evaluation, as it will vary with the considerations above and the type of data required. However, evaluation should be proportional, and it is unlikely that a small, low-cost policy will have a large, costly evaluation. It is important to bear in mind that collection of new data is costly – particularly if large-scale surveys and/or innovative methods are required.

Management and ownership: Typically, evaluations should have:

- o <u>an SRO at SCS level</u> to be responsible for the delivery of the evaluation;
- o <u>a project manager</u> to oversee the day-to-day running of the evaluation and the alignment with the policy's key evaluation questions. For externally contracted evaluations, the project manager will be responsible for commissioning, day-to-day management, ensuring the evaluation stays on track, advising the contractors and reacting to issues that develop, and essentially, ensuring the findings are fed back and used;
- o <u>suitable analytical support</u> to ensure the analytical rigour and quality of the evaluation (this role can be combined with the project manager role if the project manager is an analyst).

The level of input required will be greatest at key points (in particular, the design and commissioning stages), but throughout an evaluation there is an ongoing resource requirement that should not be underestimated.

Steering group: Establishing a steering group for the evaluation will help ensure it is designed and managed to meet the requirements of relevant stakeholders and remains on track. A steering group will usually include the policy lead(s) and relevant team members, analyst(s) supporting the evaluation, the economist responsible for the policy's appraisal, key delivery partners, other government departments, etc. Appreciating the requirements of key stakeholders and involving them throughout the evaluation will build commitment to the evaluation and help ensure that results are viewed as valid and useful. It may also be appropriate to seek co-funding from other stakeholders, if the evaluation is meeting dual evidence needs.

Steering Groups are particularly important for large-scale evaluations, but this type of scrutiny and support is always useful, even for light touch evaluations. Existing governance groups could be utilised in these cases.

Delivery bodies: A successful evaluation will depend on the engagement and cooperation of those organisations and individuals involved in delivering the policy. It will be important to work with delivery stakeholders to ensure they are aware and signed up to what the evaluation seeks to address, how, and what input will be required from them, and how they can potentially benefit from the findings, for example, by sharing data or taking part in interviews. As part of the evaluation plan, you should consider the burden being put on stakeholders, and how this can be minimised.

Analytical support: Due to the multi-disciplinary nature of evaluations it is important to consider the range of internal analytical specialists who might need to be called upon for advice and to quality assure the evaluation approach and outputs. Analytical input will be particularly important at key points: when the specification is developed and at the tendering stage, to help ensure the evaluation design is robust, and appropriately skilled contractors are commissioned; in the steering of the project; and to quality assure interim and final reports and other outputs.

The Policy Evaluation Team are available to provide support and guidance at any point, and will particularly help ensure quality assurance at key stages. For all externally commissioned evaluations the evaluation team are required to review the **evaluation justification** as part of the procurement business case approval, and sign off the **evaluation design** set out in the specification to ensure consistency and quality. Additionally, the evaluation team are keen to sit on commissioning panels for external evaluations, and steering groups for larger evaluations. Finally, to ensure quality assurance and consistency of **evaluation findings**, the evaluation team will sign-off all evaluation reports before being agreed and/or sent to ministers.

Peer review: It is good practice to have final reports independently peer reviewed, to assess the quality of the evaluation and the validity of the conclusions made and to demonstrate transparency and rigour. In some cases, particularly for complex or controversial policies/evaluations, peer review is also advisable at the design stage to ensure the best quality evaluation design is being used and to demonstrate openness and transparency. The Evaluation Team can provide further guidance on peer review.

Wider stakeholders: Depending on the scope of the evaluation, it may also involve other stakeholders – for example, people and organisations directly or indirectly affected by the policy. Again, it will be important to think about who needs to be involved, when and how.

The questions in table 6 will help identify the required resources.

Table 6: Questions to help identify resources for the evaluation

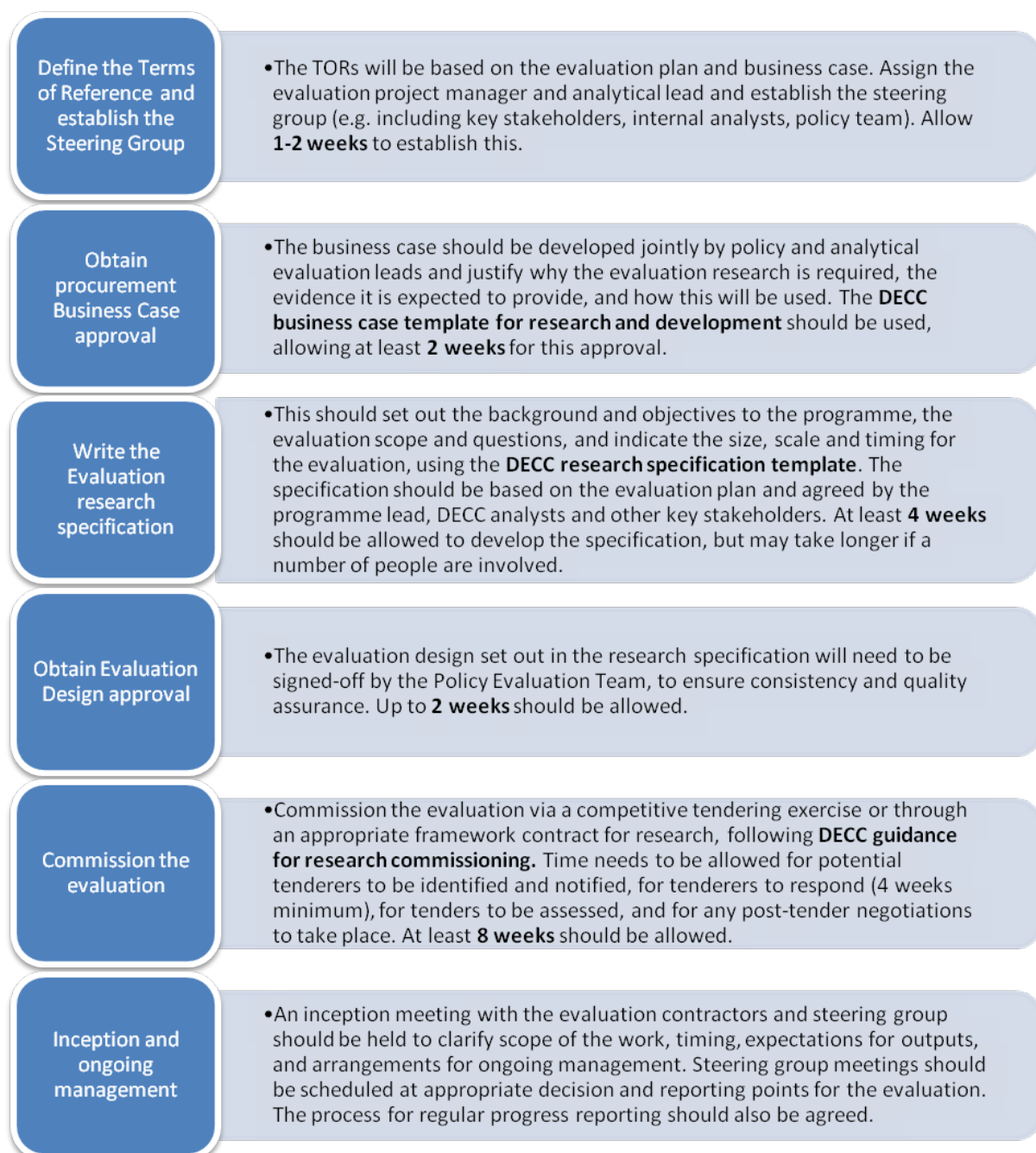- What budget is available for the evaluation? Can it be co-funded?
- Who will be the SRO for the evaluation?
- Who will be the project manager?
- How will the project be steered?
- What time and resources will be available to the project manager?
- Who will be invited to join the steering group?
- Who will provide analytical support?
- Who are the key delivery stakeholders?
- Who are the wider stakeholders?

7. Conduct / commission the evaluation

As highlighted in the previous step, the evaluation should be managed by a dedicated internal project manager. The next step is to consider exactly how the evaluation will be conducted and/or commissioned. The evaluation should be built into the policy Business Case and Delivery Plan, normally as a distinct workstream.

The flow-chart in figure 4 gives an overview of the steps for an externally commissioned evaluation, together with indicative timing for each step. An internally conducted evaluation should follow similar steps of having defined terms of reference and a project specification.

Figure 4: Commissioning and managing an evaluation

Define the Terms of Reference and establish the Steering Group	• The TORs will be based on the evaluation plan and business case. Assign the evaluation project manager and analytical lead and establish the steering group (e.g. including key stakeholders, internal analysts, policy team). Allow **1-2 weeks** to establish this.
Obtain procurement Business Case approval	• The business case should be developed jointly by policy and analytical evaluation leads and justify why the evaluation research is required, the evidence it is expected to provide, and how this will be used. The **DECC business case template for research and development** should be used, allowing at least **2 weeks** for this approval.
Write the Evaluation research specification	• This should set out the background and objectives to the programme, the evaluation scope and questions, and indicate the size, scale and timing for the evaluation, using the **DECC research specification template**. The specification should be based on the evaluation plan and agreed by the programme lead, DECC analysts and other key stakeholders. At least **4 weeks** should be allowed to develop the specification, but may take longer if a number of people are involved.
Obtain Evaluation Design approval	• The evaluation design set out in the research specification will need to be signed-off by the Policy Evaluation Team, to ensure consistency and quality assurance. Up to **2 weeks** should be allowed.
Commission the evaluation	• Commission the evaluation via a competitive tendering exercise or through an appropriate framework contract for research, following **DECC guidance for research commissioning.** Time needs to be allowed for potential tenderers to be identified and notified, for tenderers to respond (4 weeks minimum), for tenders to be assessed, and for any post-tender negotiations to take place. At least **8 weeks** should be allowed.
Inception and ongoing management	• An inception meeting with the evaluation contractors and steering group should be held to clarify scope of the work, timing, expectations for outputs, and arrangements for ongoing management. Steering group meetings should be scheduled at appropriate decision and reporting points for the evaluation. The process for regular progress reporting should also be agreed.

It will be useful to give an indication in the evaluation plan of when each of the preceding steps expect to be undertaken. See **Further Information** section for links to templates and further guidance.

The questions in table 7 will help you to think about the commissioning process.

Table 7: Questions to help clarify commissioning the evaluation

- Which aspects of the evaluation will be commissioned externally?
- Who will be involved in developing the specification?
- When will the specification be developed?
- How and when will the evaluation be commissioned?
- Who will be invited to tender?
- Who will be involved in assessing the tenders?
- When will the evaluation start?

EXAMPLE STEP 7: Securing the resources and conducting the evaluation

In the hypothetical example, the evaluation is **built into the policy business case** and programme delivery plan and has an SCS Senior Responsible Owner. A dedicated project manager in the policy team is responsible for the evaluation, with ongoing support from the analytical lead (economist) for domestic energy efficiency policy. An evaluation steering group is established which includes the key policy and analytical leads, a member of the policy evaluation team and representatives from Ofgem. Wider stakeholders including the Energy Retail Association (representing the energy companies) are consulted at key stages.

The scope and scale of the policy, and the range of evaluation questions identified, mean that an external evaluation is required. A procurement business case (developed by the project manager) is approved by the policy SRO, DECC Finance and the Head of Policy Evaluation. The project manager develops the evaluation specification, with support from the analytical lead and other steering group members. This is signed off by the Policy Evaluation Team. The tender evaluation panel for the work comprises of the project manager, the analytical lead and a member of the policy evaluation team. Day to day management of the external evaluation contractors is led by the project manager, with regular steering group meetings to ensure the work reflects policy requirements and direction.

8. Use the evaluation findings

At the time of planning an evaluation it is a good idea to give some thought to how the findings will be used and disseminated. Good practice recommends that research reports are published and accessible, but other ways of disseminating and using the evaluation evidence should also be considered early and reviewed regularly by the steering group.

Normally, evaluation findings are published in full, although related policy recommendations may not necessarily be. This ensures they have maximum impact and allows external scrutiny and challenge, as well as demonstrating to external stakeholders that policies are being robustly assessed.

As well as taking into account the publication process it is important to consider how findings will be presented. For example whether there will be a single report or a series of related reports, whether a technical report will be necessary, whether and how the related data can be made available for further scrutiny, whether presentations / seminars or other face-to-face methods could usefully be used (this is especially important to consider early if it is expected that external contractors will be expected to lead these events, as this will need to be built into their contract). The **DECC report template and report writing guidance for evaluations and social research** (in development) should normally be used. If you are commissioning the evaluation externally you will need to specify the format of the report and any presentations at the time of commissioning.

The draft and final report of evaluation findings will need to be assessed and **signed off** by the evaluation steering group and the Policy Evaluation Team, and external peer review would normally also be expected. These checks ensure that the analytical work is robust, that any recommendations or conclusions do not go beyond what can be justified given the findings and ensures that key stakeholders have been able to have sight of the conclusions and comment on any factual inaccuracies.

EXAMPLE STEP 8: Using the evaluation findings

Following the early roll-out stage of policy, the programme board receive quarterly updates of emerging impact findings from the evaluation, including measures installed and assessment of behaviour change across different delivery models. Ongoing process evaluation allows these differences in uptake and early impacts to be better understood, and also identifies some unintended outcomes of the policy. A workshop takes place at the end of the early roll-out stage, bringing together the evaluators with policy leads to discuss the emerging findings. This is then followed by a similar event jointly led by evaluator and policy for business stakeholders to present emerging findings and conclusions and subsequent publication of interim findings. This information is used to amend certain aspects of the policy at the first review point.

The findings of the process evaluation, along with findings about the overall impact of the policy are presented in a final evaluation report at the completion of policy delivery. The findings and are used to inform a new policy and delivery model, and the underlying data feeds into the impact assessment and modelling for this.

The **Policy Evaluation Team** and other analysts within the Department (economists, social researchers and statisticians who support policy teams) will be able to offer further advice on monitoring and evaluation issues, and should be contacted in the first instance. They will also be able to advise on how the evaluation relates to impact assessments and Post-Implementation Review (PIR).

Guidance and templates for research procurement and reporting are available on the following DECC intranet pages:

DECC Research commissioning guidance and Business Case template

Evaluation and social research Invitation To Tender template (under development)

DECC report template

Report writing guidance for evaluations and social research (under development)

The key government guidance on policy evaluation is the HM Treasury **Magenta Book**. Part A provides an overview for policy-makers and sets out what evaluation is, and what the benefits of good evaluation are. It explains in simple terms the requirements for good evaluation, and some simple steps that policy-makers can take to make a good evaluation of their intervention more feasible. Section B is designed as more of a specialist resource for analysts and provides technical guidance on designing and undertaking evaluations.

Magenta Book: Section A (under development, to be published March 2011)

Magenta Book: Section B (under development, to be published summer 2011)

The Inter-departmental Analysts Group (IAG), chaired by DECC, provide advice on methodological approaches for estimating and valuing climate and energy impacts of policies. **Guidance on valuing greenhouse gas emissions** is available at: http://www.decc.gov.uk/en/content/cms/statistics/analysts_group/analysts_group.aspx

The Better Regulation Executive has guidance available outlining the requirements for conducting **Impact Assessment Post-Implementation Review**. This can be found at: http://www.bis.gov.uk/assets/biscore/better-regulation/docs/10-900-impact-assessment-quick-start-guide.pdf

The Department for Transport has produced guidance on **developing logic models** which, includes useful advice transferable to a range of policy contexts. This is available at: http://www.dft.gov.uk/pgr/evaluation/evaluationguidance/logicmapping/