

Annex A: guidance for when and how to evaluate

This annex contains more detail on the criteria FSA colleagues should consider when deciding whether to evaluate their work, how to decide the scale of an evaluation and how to decide what type of evaluation is required.

Evaluation criteria

The FSA has developed a set of criteria based on best practice guidance that should be considered when deciding whether and how to evaluate FSA activities. These are:

- whether there is a knowledge gap to be filled by evaluation, including when the policy or strategy being implemented is novel or untested or where evaluation evidence would inform future practice
- whether evaluation is required to demonstrate accountability and transparency over use of public funds and the fitness for purpose of agency interventions
- what the scale of investment associated with the policy, programme or project being evaluated has been
- whether conducting an evaluation would be feasible and deliver evidence in a timely manner

These criteria are intended to be used alongside and to complement existing tools, most notably the FSA's business case process. This process requires colleagues to articulate the anticipated benefit of their proposed activity, current performance in the area (for example, baseline data) and anticipated measures.

Evaluation scale

Activities can be evaluated in many ways. The nature and range of work done by the FSA, which commonly involves a range of delivery partners as well as cross-organisation and cross-nation working, means a tailored approach to evaluation is required.

Although final decisions on the scale of evaluation required will have to be taken in the context of wider business needs, available resources, organisational priorities, and what has been specified within business cases, Figure 2 provides a rule-of-thumb guide which FSA colleagues can follow to help choose an appropriate scale of an evaluation. This framework is based on that developed by the [UK Space Agency](#).

Figure 2: choosing proportionate evaluation

Risk and uncertainty:

- Low - Straightforward low-risk programme with low uncertainty around the outcomes
- Medium - Programme not especially complex or risky, but some uncertainty around outcomes

- High - Complex programme design, and/or significant risk and uncertainty around programme outcomes

Budget and Profile:

- High - Large programme with significant budget, and/or high profile with public interest, and potentially high impact
- Medium - Medium-sized programme with moderate budget, and/or some public interest, expected to have a sizeable impact
- Low - Small budget and/or limited public interest

Category	Risk and uncertainty: Low	Risk and uncertainty: Medium	Risk and uncertainty: High
Budget and Profile: High	Level 2	Level 3	Level 3
Budget and Profile: Medium	Level 2	Level 2	Level 3
Budget and Profile: Low	Level 1	Level 2	Level 2

Level 1: light-touch evaluation recommended, including before/after monitoring

Level 2: consider commissioning externally, with appropriate budget allocation

Level 3: detailed, externally commissioned evaluation with budget of 1-5% of total programme recommended

* Budget thresholds: < £150,000 Low; £150,001-£500,000 Medium; £500,001+ High

The framework recommends that FSA colleagues consider the budget and profile of the activities being evaluated alongside the risk associated with the work and uncertainty over what outcomes and evaluation would deliver when deciding what level of evaluation is needed.

Activities which are lower risk and where there is low uncertainty around the outcomes of an evaluation and where budgets are small would likely suit a 'light-touch' evaluation that could include before / after monitoring (Level 1).

Activities where there is some uncertainty around outcomes, but which are not especially complex, may require a larger-scale evaluation, with appropriate budget2 allocation (Level 2). These activities may require a Level 3 evaluation – a detailed, externally commissioned

evaluation with appropriate budget (1-5% of total programme recommended) - when the activity being evaluated has a significant budget, and/or there is high interest in the outcome, and potentially high impact in conducting an evaluation for the future of the policy or programme.

Activities that have complex programme designs, and/or significant risk and uncertainty around programme outcomes typically require a Level 3 evaluation, although in situations where the activities have small budgets or where there is limited media/public interest a Level 2 evaluation may be suitable.

Levels do not need to be distinct: there could be scope for including elements from two levels, for example. Likewise, not all Level 2 or Level 3 evaluations will require externally commissioned evaluations. In some instances, internally led evaluations may be more appropriate and represent a more effective use of resources. Similarly, while benefits measurement may be delivered through evidence generated in-house, externally commissioned evaluation may provide evidence of unanticipated consequences of an intervention and a more nuanced understanding of the effect of activities.

We anticipate that evaluation of the FSA's priority programmes and corporate priorities will require either Level 2 or Level 3 evaluations, but that constituent parts of these programmes may require Level 1 or Level 2 evaluations.

Evaluation type

There are three common types of evaluation: process, impact and value-for-money.

1. Process evaluations consider whether an intervention is being implemented / delivered as intended, whether the design is working and what is working more or less effectively, for whom and why.
2. Impact evaluations involve an objective test of what changes have occurred, the scale of those changes in an assessment of the extent to which they can be attributed to the intervention.
3. Value-for-money evaluation involve comparing the benefits and costs of the intervention.

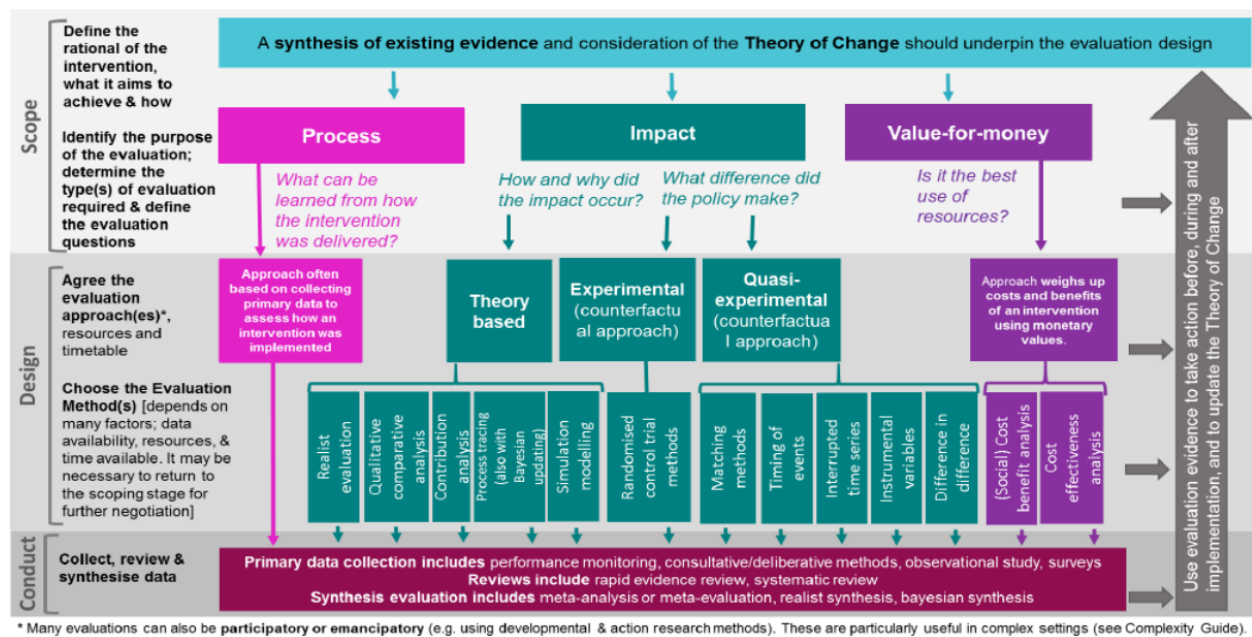
Which type of evaluation is appropriate depends on the questions being addressed by the evaluation. While further details are available in the Magenta Book, process evaluations typically seek to understand what can be learned from how the intervention was delivered, impact evaluations try to understand what difference the intervention made, while value-for money evaluations seek to address whether the intervention is a good use of resources.

The [Magenta Book](#) recommends conducting scoping work prior to deciding the type of evaluation required. For this to be done effectively, a broad range of internal, and sometimes external, stakeholders need to be engaged. In line with the ROAMEF Policy Development Cycle, this scoping work should take place alongside policy development and prior to implementation. It is essential that evaluation be considered at the start of the policy development and implementation process; suitable benefit measures and plans for how these can be realised must be included in business cases. This is because how the policy is implemented and what data is collected during implementation holds implications for what types of evaluation are feasible. This is particularly the case for impact evaluations, where control/comparison groups are often necessary to demonstrate impact.

Process, impact and value for money evaluations require different approaches and resources (see Figure 3). Specific guidance on this is available in the Magenta Book Annex A. This details the [analytical methods for use within an evaluation](#), including generic research methods for use in process and impact evaluations, methods for experimental and quasi experimental methods for impact evaluation, theory based methods, methods for value for money evaluation and methods

for the synthesis of existing evidence. Regardless of evaluation type, most evaluations benefit from quantitative methods being included in some capacity.

Figure 3: scoping, designing and conducting an evaluation from the Magenta Book



The type of evaluation approach and methods used should always be informed by the research questions being addressed and considerations of what is feasible in a given context. While best practice guidance may recommend experimental approaches that can demonstrate causal relationships between control and comparison groups (for example, [Level 3 in The NESTA Standards of Evidence](#)), and such experimental methods are considered by default by the FSA, this may not always be practicable within an organisation's structure or its regulatory responsibilities. Likewise, certain evaluation types can only be conducted if appropriate baseline data is collected. It is therefore essential that evaluation is not an afterthought but instead integrated into the policy development process.