SCOTTISH ENTERPRISE

Evaluation Guidance Note

A Summary Guide to Evaluating Economic Development Projects or Programmes

Appraisal & Evaluation Team
1st November 2008

Scottish Enterprise, 150 Broomielaw, Atlantic Quay, Glasgow, G2 8LU

RECORD OF CHANGES

Version	Change Date	Description
v1.0	1/11/2008	
V1.1	10/11/2011	Hyperlink updates

CONTENTS

1 INTRODUCTION

1.1 PURPOSE OF THIS GUIDANCE NOTE

This note provides a summary description of how to evaluate an economic development project or programme.

The aim of this note is to promote consistent, good quality assessment of the delivery and economic benefits from Scottish Enterprise interventions, and to assist Scottish Enterprise staff and contractors in identifying and using a good practice approach.

Without evaluation we do not know if the intervention is delivering results against the objectives that have been set, nor can we tell if the intervention offers good value for money. There is a duty on Scottish Enterprise to make the best use of the available resources and to demonstrate to stakeholders that decision making is based on the best available evidence.

This note draws on and adapts other guidance (see list of sources in section 13.1).

1.2 RELATIONSHIP TO OTHER GUIDANCE

This supersedes all other previous Scottish Enterprise guidance on evaluation.

This guidance note is also consistent with the high level discussion of principles and best practice in project appraisal and evaluation as presented in HM Treasury (2003) The Green Book, Appraisal and Evaluation in Central Government.

1.3 WHAT IS DIFFERENT IN THIS GUIDANCE?

There are number of differences between the approach to evaluation in this guidance and previous Scottish Enterprise guidance. These differences include:

Scope: this guidance note introduces, to a basic level, a range of research and evaluation topics. The material is intended to assist non-specialists in undertaking relatively straightforward evaluation activities, and also, to allow them to act as more informed customers, when using external consultancy support.

Expectations on Evaluation Activities: There is an increased demand for evidence based decision-making to inform economic development choices and priorities. While recognising that evaluation activity must be proportionate to the intervention being considered, there are expectations of higher and explicit standards in evaluation research design, methods, delivery, analysis and reporting.

Organisational Context: Scottish Enterprise has changed in terms of the focus of economic development support and the organisational structures to support delivery. These changes present new challenges for evaluation, such as an increased focus on the evaluation of innovation activity and an organisation-wide evaluation focus (rather than Local Enterprise Company). Local office specialist evaluation capacity has also been pooled into a single Appraisal & Evaluation team. These changes place an increased emphasis on clear and consistent guidance to support evaluation activity.

The Context of Project and Programme Evaluation: There is an ongoing commitment within Scottish Enterprise to ensure consistent approaches to evaluation and a direct link between the evidence base and policy development and implementation. This is being addressed by the development of Policy Evaluation Frameworks (PEFs) which represent a coordinated approach to evaluation, research, strategic direction setting, and consideration of future policy design and implementation.

Capacity Building: It is anticipated that effective implementation of the material in this guidance will require, depending on user experience, varying levels of training, mentoring and shadowing on a regular basis.

Checklists and 'Hints & Tips': a range of checklists and 'hints & tips' are included to aid consistent and effective evaluation activity.

1.4 THE ORGANISATION OF EVALUATION IN SCOTTISH ENTERPRISE

All evaluation in Scottish Enterprise are co-ordinated by the Strategic Evaluation and Research Group (SEER Group). This group, comprised of representatives from across different business units, ensures there is a single organisational approach to developing the evidence base and ensuring a strategic approach to evaluation and research.

2 WHY DO EVALUATION?

2.1 THE RATIONALE FOR EVALUATION

The first question that needs to be asked when planning or starting an evaluation is:

- **For what reasons are we undertaking the evaluation?**

The rationale for evaluation is set out by HM Treasury as follows:

All new policies, programmes and projects, whether revenue, capital or regulatory, should be subject to comprehensive but proportionate assessment, wherever it is practicable, so as best to promote the public interest. The Green Book presents the techniques and issues that should be considered when carrying out assessments (HM Treasury 2003, p1).

The HM Treasury Green Book guidance, also notes the following reasons for evaluation:

- to explore what is working well and what is not in a project, programme or policy,
    - i.e. learning and best practice (including management, content, delivery, recruitment and how far the intervention is meeting beneficiaries' needs) in order to decide whether the intervention should be continued in its current form, altered and improved, expanded or discontinued; and
- To establish the outcomes and impacts of an intervention.

Investing time, money and effort in evaluation has to be justified in terms of the difference it makes to policy success.

Evaluation is not an end in itself. In economic development the policy concern is to enhance the economic (and sometimes social) prospects of individuals, regions or sectors.

The justification for evaluation in all these cases is the same: **can we apply evaluation procedures and methods in ways that will improve the quality of life, prosperity and opportunities available to Scotland's citizens?**

To make a difference in this way requires that evaluation asks and answers questions that are useful to policy stakeholders; whether they are project owners, managers, policy makers or beneficiaries.

There are a number of particular contributions that evaluation can make throughout the programme cycle, as illustrated in the figure below.

Figure 2.1 Contributions of Evaluation

Designing programmes

- One of the core competencies of evaluation is to gather information from different stakeholders, which can ensure the relevance of projects to the needs of users . There may be a continuous process of feedback from users and from other stakeholders throughout a project.

Choosing between interventions

- A well-designed evaluation system, as is the case for appraisal, will also contribute to the selection of specific interventions within the general scope of a policy area.

Improving management and delivery

- A fully integrated evaluation process can make a key contribution to the way projects are managed and delivered.

Identifying outputs, outcomes and impacts

- From an early stage, interventions need to demonstrate results. For many policy makers, identifying, describing and quantifying such outputs, outcomes and impacts is a major benefit of evaluation.

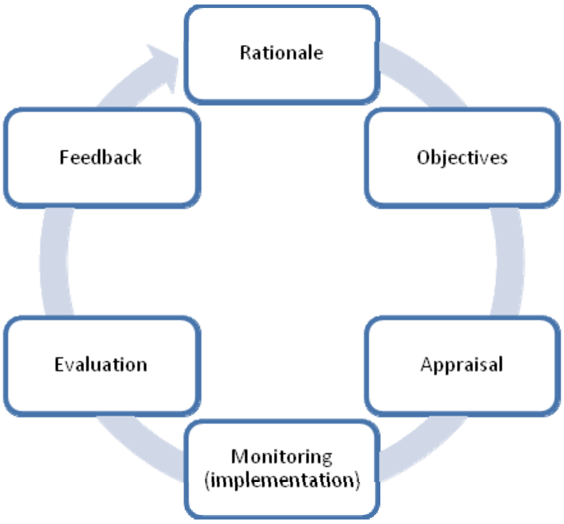
Identifying unintended consequences and 'perverse' effects

- Even when projects fulfil their stated objectives there will also often be unintended consequences. These can be positive or negative and should be assessed.

2.2 WHERE DOES EVALUATION FIT IN THE LIFE OF A PROJECT OR PROGRAMME?

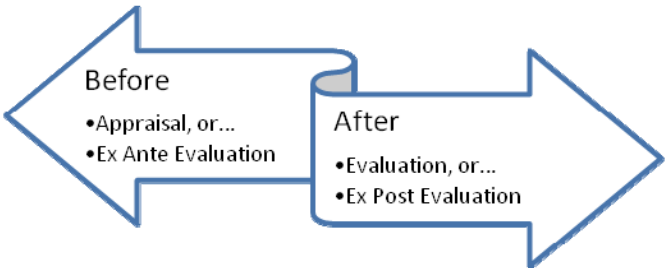
Evaluation often forms a stage in a broad project, programme or policy cycle. This is sometimes called the ROAMEF cycle (standing for **R**ationale, **O**bjectives, **A**ppraisal, **M**onitoring, **E**valuation and **F**eedback).

Figure 2.2 ROAMEF Diagram



Evaluation represents an assessment of an intervention after resources have been committed. This is also sometimes referred to as **Ex Post Evaluation**.

Figure 2.3 Evaluation Terminology



Levels of Evaluation: Project, Programme, or Policy?

The problem of linking policies, programmes and specific interventions, or projects, is a perennial one in evaluation.

There is now a tendency for evaluation to move 'upstream' and pay increasing attention to the policy level. This is also the case at Scottish Enterprise, where **Policy Evaluation Frameworks** are being established for each main business unit.

This is good. This reflects a willingness of policy makers to take on board evaluation results.

At the same time it presents **challenges** for evaluators who need to view the results of their work in a wider context. Considering the policy level can also strengthen project evaluation, for example by identifying results oriented criteria for project success.

Users of this guidance note should consult with the members of their business unit involved in the Policy Evaluation Framework for that area, or with the Appraisal and Evaluation Team, in order to ensure any specific evaluation activity is properly integrated into Policy Evaluation Framework plans.

2.3 DIFFERENT APPROACHES TO EVALUATION

Evaluation has varied roots, is derived from different research traditions, and informed by different sets of ideas. It follows from this, that there is no one right way to do evaluation. Specialist evaluators themselves often have different priorities depending on their background: they may be economists concerned with efficiency and costs; management consultants interested in the smooth running of the organisation; policy analysts with a commitment to public sector reform and transparency; or scientists (of various disciplines) concerned to establish truth, generate new knowledge and test research hypotheses.

To take this further it is useful to be generally aware that there are different theoretical approaches that often inform approaches evaluation. This guide does not consider these in depth, but it may be helpful to be aware of some of the terminology (see Figure 2.4). Despite these differences in evaluation origins and traditions, it is possible to distinguish some of the **main types of evaluation**. Focusing on the purposes of evaluation, the most commonly recognised purposes are shown in Figure 2.5 below.

It is normal for Scottish Enterprise evaluations to reflect a range of these approaches and purposes, but often varying by the timing and specific remit of an evaluation. It is important to emphasise however, that in Scottish Enterprise evaluation, estimation of **economic impact** normally forms a **central focus**, specifically estimations of Gross Value Added as a primary metric, firmly located within the Government Economic Strategy framework.

Figure 2.4 Approaches to Evaluation

Programme theory:

- The elaboration of the **Logic Models**, used extensively in the context of Dept for International Development, World Bank and EU funded development programmes is one kind of simple evaluation theory that focuses on programme inputs, outputs, results and impacts.
- The **Theory of Change** is a programme theory approach that is concerned with opening up the black box and going beyond input output descriptions and seeking to understand what are the 'theories' of actors with regard to programme interventions and why they should work.

Theories about the practice of evaluation:

- There is a growing literature on evaluation practice, ie what evaluation attempts to do and what appear to be effective approaches.
- Such theories are the distillation of what has been learned through studies of past evaluations.
- For example, how to ensure that evaluations are used, how to draw conclusions from evidence, and how to put a value on a programme or intervention.

Theories of implementation and change:

- These include understandings of: policy change; the diffusion of innovation; administrative and organisational behaviour and leadership.
- The theories are mainly derived from political science and organisational studies. Their application in evaluation may condition the success of programme interventions.

Policy specific theories:

- There is a body of theory associated with socio-economic development, ie how does development occur spatially and sectorally.
- There are similar bodies of theory in most policy priority areas eg in education; health; employment; environmental protection etc.
- Because the design of interventions is usually underpinned by a rationale that derives from theory within policy areas, it is both useful and normal that evaluators have some knowledge of the theories relevant to the themes and policy areas under consideration.

Figure 2.5 Main Purposes of Evaluation

Planning/efficiency

- Ensuring that there is a justification for a policy/programme /project and that resources are efficiently deployed.

Accountability

- Demonstrating how far a programme/project has achieved its objectives and how well it has used its resources.

Implementation

- Improving the performance of programmes / projects and the effectiveness of how they are delivered and managed.

Knowledge production

- Increasing our understanding of what works in what circumstances and how different measures and interventions can be made more effective.

Institutional strengthening

- Improving and developing capacity among programme/ project participants and their networks and institutions.

3 PLANNING & STRUCTURING EVALUATION

This section discusses the various activities and issues that are needed when planning evaluation work.

3.1 ESTABLISHING THE AIMS AND OBJECTIVES OF EVALUATIONS

Defining the scope of an evaluation amounts to asking the question:

- **What is going to be evaluated?**

This is a vital part of planning any evaluation because it would be an endless task to look at every aspect of a project or programme. When answering this question, it will be important to consider which of the project/programme processes and other evaluation issues need to explored and which need to be explored in more detail than others. It is normally helpful to adopt a relatively strict definition of the scope of the evaluation.

What can be achieved in the evaluation will be influenced by the goal and by the availability of data, time constraints and limitations on financial resources. In order to reach conclusions, the evaluation should be confined to an examination of the programme and its most essential interdependencies with other public policies and interventions.

3.2 KEY QUESTIONS FOR SCOTTISH ENTERPRISE EVALUATION

A list of the typical issues for evaluation to focus on within Scottish Enterprise is noted below in Figure 3.1. It is important to emphasise however, that in Scottish Enterprise evaluation, estimation of **economic impact** often forms a **central focus**, specifically estimations of Gross Value Added as a primary metric, firmly located within the Government Economic Strategy framework. In addition, it is noted that not all of these issues will need to be addressed every time. For a number of these topics, particular issues are highlighted for consideration.

Figure 3.1 Evaluation Topic Areas

Inputs
- What did the project/programme cost?
- Costs should take into account all public sector costs and permit, where possible, a split between Scottish Enteprise and other costs.

Progress Towards Aims and Objectives
- Have the project/programme aims and objectives been achieved?
- Are they still relevant considering the changing context and needs of beneficiaries?

Management
- Is the project/programme management appropriate and effective?
- Can improvements be made in the management of the project/programme?
- Are partnership arrangements satisfactory? What alterations are needed?

Content and Delivery
- Was the content and delivery appropriate and effective?
- Was the target group reached and was it the right one? Are changes required?
- Is the scale of the project/programme appropriate?
- What lessons can be learned?

Progress Towards Targets, Performance Indicators, Quality Indicators and Outputs
- Were the target milestones, performance and quality indicators achieved?
- Were the target outputs achieved? Were ethnicity and other targets met?
- Is monitoring information being accurately and effectively collected, analysed, disseminated and responded to?

Outcomes
- Did the project/programme result in the intended outcomes for beneficiaries? These might be hard outcomes (e.g. entry into learning and employment) or soft outcomes (e.g. raised levels of confidence, motivation and self-esteem).
- Do the outcomes meet the needs of target beneficiaries?
- Are the outcomes sustainable?
- Is there evidence of clear added value?
- Confounding factors: have any changes in the economic environment been taken into account where they have affected performance?

Impacts
- Were the anticipated wider impacts achieved (taking into consideration issues of additionality, e.g. deadweight, displacement and multipliers/linkages)?
- Do the impacts meet the needs of target beneficiaries?
- Are the impacts sustainable?
- What progress has there been from the socio-economic baseline as a context to the project or programme's impacts?

Value for Money: effectiveness, efficiency and economy
- Is the project/programme effective, efficient (cost-effective)?, and economical?
- Is overall Value for Money being achieved?
- Where cost-benefit analysis is undertaken, discounting and calculation of Net Present Values should be undertaken

Learning and Best Practice
- What has been learnt?
- What best practice has been exemplified?

Succession
- Is there a continuing need for the project/programme activities? What is the extent of this need?
- Will the continuing need be met by other agencies or should the project/programme activities be continued? How might they be funded?

3.2.1 THE STRATEGIC & POLICY CONTEXT

The project/programme will have originally been developed within the context of appropriate strategic frameworks. These policy aims and objectives should be identified and recorded.

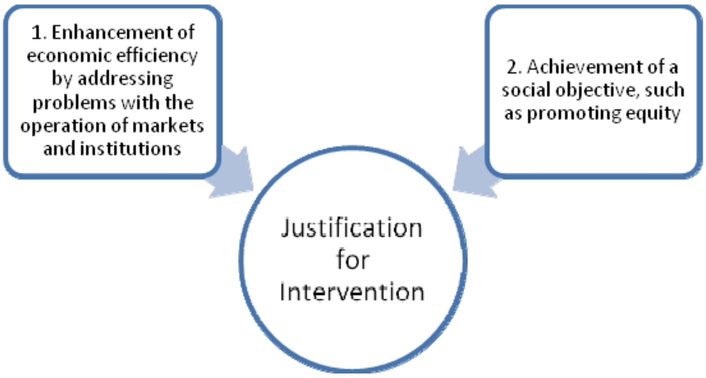
However, frameworks develop and change over time and at the evaluation stage it is important to reassess the project/programme's continued fit with and contribution to relevant strategic frameworks.

A key step is to identify the policy goals that the intervention has contributed to. This involves an assessment of Scottish Government and Scottish Enterprise policy aims and objectives.

3.2.2 ESTABLISH THE MARKET FAILURE & EQUITY RATIONALE

Economists identify two broad types of reason why government activity may be required: **Market Failure** and **Equity** as illustrated below. These issues should be considered and the extent to which they relate to your project should be explained.

Figure 3.2 Reasons for Expenditure



The aim of intervention is to address the market failure in a beneficial and sustainable way. Ideally, this means encouraging the economic actors in the relevant market to perform more effectively and in a sustainable way so that Scottish Enterprise can exit.

Key considerations for the evaluation will be:

- the extent to which the market failures have been addressed;
- whether this has resulted in a re-adjustment in the market;
- whether the intervention should continue; and
- What form any future intervention should take.

It should be noted that market failures can exist at both the level of the overall market and individual beneficiary level. For example, a project/programme may encourage an individual company to operate in a different way in the future and in doing so address the market failure at the company level. However, the project/programme may not affect the behaviour of non-supported companies and therefore does not address the market failure at the level of the overall market. The main types of market failure are indicated in the figure below.

Figure 3.3 Market Failure

Imperfect competition (market power).

- Economic theory demonstrates efficient outcomes will be delivered only where markets are actually or potentially competitive.
- As soon as there is an element of monopoly (on the side of the seller) or monopsony (on the side of the buyer) power that can be exercised, a less efficient outcome will occur.
- This may arise because of the natural characteristics of the market (e.g. very high costs of entry) or through strategic behaviour by incumbents (e.g. predatory pricing).

Externalities.

- Externalities result when a particular activity produces benefits or costs for other activities that are not directly priced into the market.
- When this happens, the amount of the particular activity that takes place will generally be inefficient. Externalities can be "positive" or "negative".
- An example of a positive externality is the spill over effect into other areas that can occur as a result of research and development activity.
- A company or research institution will generally decide its level of R&D on the basis of the benefits that it can capture - ignoring benefits that might occur elsewhere.
- An example of a negative externality is pollution of the environment.
- A company or individual may reduce its own costs by failing to implement pollution controls, but this will generally impose costs on those affected by the pollution.

Information failures.

- The effective operation of markets relies on the fact that all the participants in the market have complete and perfect information relevant to that market.
- When this information is not available to all participants, this is described as asymmetry of information, and market failure can arise.
- Information asymmetries lead to sub-optimal outcomes.
- For example, a buyer may not have full information on the characteristics of a product or service he/she wishes to buy - this is known as adverse selection.

Public goods.

- Markets work effectively to provide private goods and services, which are typically rival and excludable in nature - i.e. each specific item or service can only once be sold/bought, and once purchased, can be exclusively "enjoyed" by the purchaser.
- In contrast, public goods and services are non-rival and non-excludable - if one person purchases the good or service, that does not stop others from purchasing it; and there is generally no way to stop people from enjoying the good or service.
- True public goods and services are comparatively rare, but the provision of national defence and of law and order are typically used as illustrations.

Equity is to do with the delivery of social or distributional objectives. Even where markets are working efficiently, they may result in a distribution of income (or other benefits/costs) that is unacceptable to society. These issues should be considered where they are relevant.

3.2.3 INPUTS TO IMPACTS & ADDITIONALITY

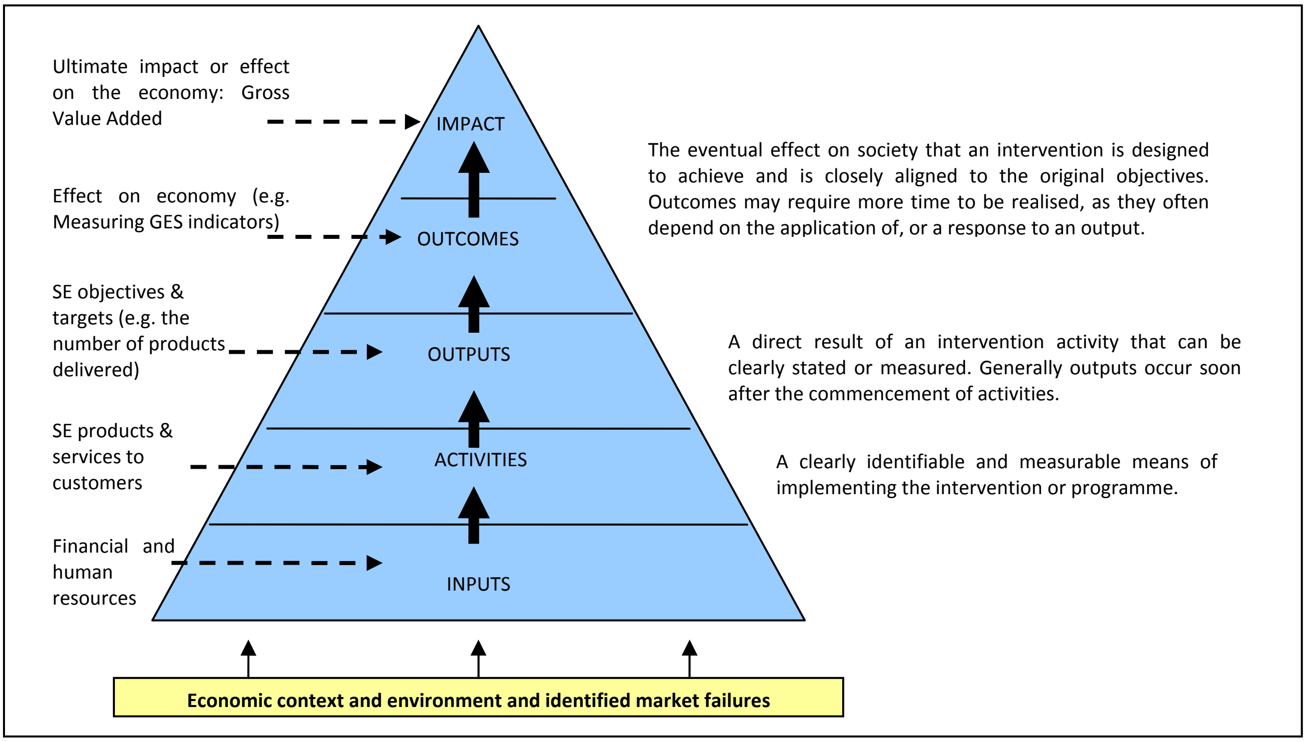
Targets should be set for project/programmes at the appraisal stage. A key role of the evaluation is to assess performance against these targets in terms of:

- the extent to which the project has met or is falling short of its targets; and
- the reason for under or over performance.

The learning from evaluation should feed back into the appraisal process. At the appraisal stage, benefits are forecast using information available at a point in time. It therefore gives an indication of anticipated benefits. An evaluation of an intervention therefore helps identify the level of any over or under optimism applied at the appraisal stage. This can help inform optimism bias (see Appraisal Guidance Note) adjustments for similar or related interventions at the appraisal stage.

In terms of assessing the route to impacts of its interventions, Scottish Enterprise adopts the inputs to impacts approach as set out in Figure 3.4 below. In implementing this process and to move from gross to net effects it is necessary to assess additionality. Additionality is the difference between the reference case position (what would have happened anyway) and the position after the intervention (intervention option) was implemented.

Figure 3.4 Elements of Economic Development Process

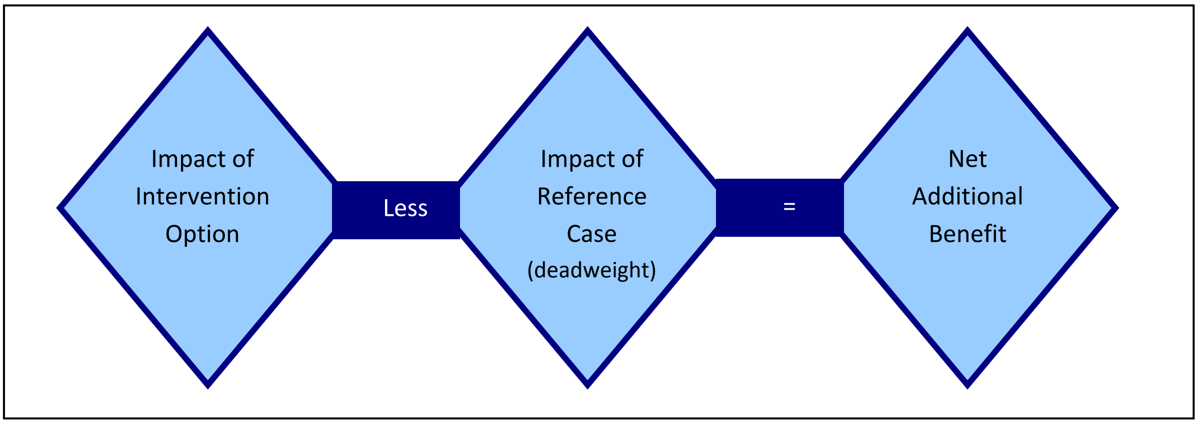


Source: Adapted from Scottish Enterprise (2005) Measuring Gross Value Added and the Impact of Activities, Strategy Directorate, 7th November, Scottish Enterprise: Glasgow p2.

Most Scottish Enterprise interventions will have both positive and negative effects. In appraising or evaluating the effects of an intervention it is important that all of these are taken into account in order to assess the additional benefit or additionality of the intervention – in other words, the net changes that are brought about over and above what would take place anyway.

The additional benefit of an intervention is the difference between the reference case position (what would happen anyway) and the position if / when the intervention (intervention option) is implemented (see Figure 3.5 below).

Figure 3.5 Summary of Additionality



Source: Adapted from English Partnerships (2004) Additionality Guide, A Standard Approach to Assessing the Additional Impacts of Projects, Method Statement, Second Edition, September, English Partnerships: London p3.

Additionality may relate to[1]:

> **Scale**: for example, a greater quantity of business turnover or jobs may be delivered in an area. Absolute additionality refers to a situation where all of the intervention's benefits may be claimed, in relation to the output or outcome being examined.
>
> **Timing**: for example, an activity may happen earlier than would otherwise have been the case.
>
> **Quality**: where the benefits of an intervention cannot easily be valued, then the quality of the benefits may be different because of an SE intervention, for example, the skill level of employment outcomes may be enhanced.

Scale additionality is the most significant type when it comes to assessing overall economic impact in terms of Gross Value Added (GVA).

However, consideration should also be given to other types of additionality if they have particular significance to a given intervention, for example, if the relevant benefits would otherwise have been significantly delayed, or would have been of inferior quality.

Additionality is relevant to all of an intervention's intended outputs, outcomes and impacts. These will normally be set out in the intervention proposal or approval paper, and may include a range of performance measures.

While many projects are concerned with delivering broad outcomes and impacts, the focus of attention in project appraisals and evaluations is often on assessing the net additionality of specific *outputs.* Common outcomes are employment and turnover improvements. Assessment of additionality may also be relevant to other intended outputs and outcomes.

Many outcomes (such as changing behaviours and attitudes) are difficult to measure and will often only occur sometime after an intervention has been implemented. Nonetheless, it is important that these benefits are assessed and reflected in judgements and decision-making based on appraisals and evaluations and quantified wherever possible.

The resources allocated to assessing the additional benefit of an intervention should be proportionate to the nature and scale of the intervention and the particular phase of the intervention.

[1] Technical Note: Additionality as it relates to an area or group is not distinguished here as a separate type of additionality. The effect of intervention benefits on a specific area or group is taken here to be addressed by the concept of leakage.

Generally speaking, an assessment of additionality will be required at the appraisal stage, before commencing an intervention, and when conducting full impact evaluations, after the end of an intervention or after it has been running for several years. A detailed assessment of additionality is generally not required for early intervention assessments, or interim evaluations, which occur closer to the intervention start date, and are usually more focused on ensuring the intervention is on track, and achieving immediate objectives.

3.2.4 WHAT TIME PERIOD SHOULD EVALUATION COVER?

The time period for full impact evaluation should relate to the life of the intervention, capturing all the important cost and benefits over the period (reporting the annual and/or cumulative benefits), and where appropriate, any significant anticipated benefits following the end of the formal intervention period (e.g. 1-3 years after).

Timing of the evaluation is an important consideration when assessing the benefits generated. It will take longer for some benefits to materialise. The earlier in the process the evaluation is undertaken the fewer of the benefits it will be possible to measure. In these cases it may be appropriate to assess results achieved to date and potential future (depending on the intervention).

3.2.5 GROSS VALUE ADDED

Project managers are being asked to assess the contribution to the economy of their products and projects at either the appraisal or evaluation stage. This should consider a range of measures, the **key** one of which should be, where possible, Gross Value Added (GVA): this is the ultimate effect on the economy. This should be measured as an increase in GVA as a result of outcomes.

Gross value added (GVA) (also known as GDP at market prices) is an indicator of wealth creation and measures the contribution to the economy of each individual producer, industry or sector in Scotland. GVA is generally regarded as the best measure of the sum of economic activity within an area.

GVA is the income generated by organisations out of which is paid wages and salaries, the cost of capital investment and financial charges, before arriving at a figure for profit. GVA includes taxes on production (net of subsidies), such as business rates but excludes subsidies and taxes on products (e.g. VAT and excise duty).

GVA = Employee Costs + Depreciation + Amortisation + Operating Profit + (taxes on production - subsidies on production)

Gross and net additional GVA should be identified – i.e. total GVA and GVA which would not have been achieved without the intervention of Scottish Enterprise (net additional GVA taking into account of arrange of additionality factors.

3.3 WHO IS THE EVALUATION FOR? INDENTIFYING AND INVOLVING STAKEHOLDERS

The **first question that must be asked**, after the scope of the evaluation has been defined is:

- Who are the individuals, the groups, or the organisations that have an interest in the intervention to be evaluated and/or are interested in the process or in the results of the evaluation itself?

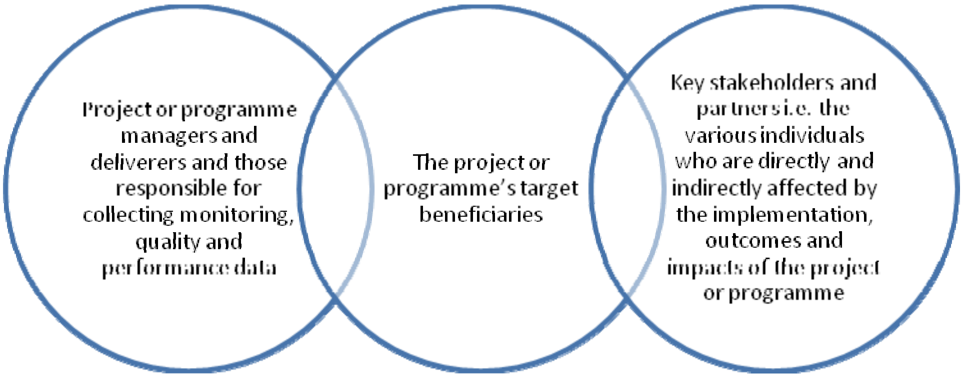
A range of different individuals and groups will need to be involved in the evaluation, although who is consulted will vary depending on the nature, size, scope and importance of the project or programme, the stage it is at and the type of evaluation being undertaken.

Ideally this exercise should take place before defining the details of the evaluation to be performed: by taking into consideration their points of view, it is possible to decide the most relevant questions that should be answered. But even if this is not possible – for instance because it has not been possible to identify all the interested parties at an early stage – some sort of involvement is desirable.

It is worth adding a point that stakeholders will have their own interests and that the independence and objectivity of the evaluation research is crucial: some stakeholders may be 'voiceless' e.g. businesses in general, taxpayers etc.

Most evaluations will include consultations with the group shown in Figure 3.6 below.

Figure 3.6 Stakeholders



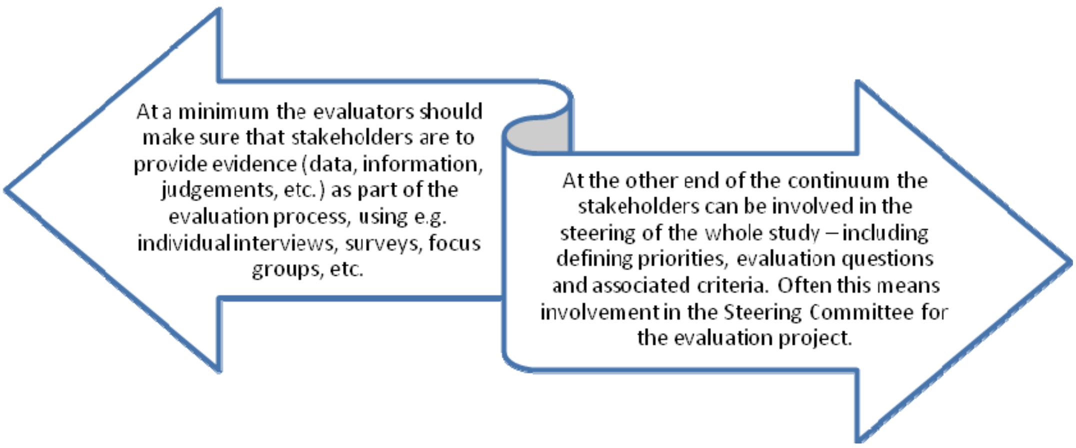
The **second question that should be asked is**:

- How is it possible to make sure that the stakeholders provide the relevant inputs to the design, management or content of the evaluative exercise?

The involvement of the stakeholders can take place at very different levels as illustrated in Figure 3.7. In practice the involvement of the stakeholders in most programmes falls somewhere in the middle.

If the participation of stakeholders in the Steering Committee is to be restricted to the official partners (e.g. Scottish Enterprise, Scottish Government, Local Government), some way to provide feedback to other actors that are able to provide information and judgements is widely practised through, for example, the dissemination of reports, and ad hoc meetings.

Figure 3.7 Involvement of Stakeholders



3.4 DRAWING UP AN EVALUATION PLAN AND TERMS OF REFERENCE

These two are similar but for clarification: The **Evaluation Plan** is an internal document and the **Terms of Reference** is a document sent out to external consultants inviting them to tender to undertake the evaluation (also known as Evaluation Specification or Evaluation Brief).

The Terms of Reference is the document that serves as the main basis of a contractual relationship between the commissioner of an evaluation and the team responsible for carrying out the work.

Devising the Terms of Reference is clearly a vital step when an evaluation has to be performed by outside consultants. This work is equally important when part of the evaluation is performed in-house.

The Terms of Reference may concern either the evaluation operation as a whole (when it is entrusted to a single team) or a part of the research work programmed in the evaluation project (in-depth analysis of an evaluative question).

3.4.1 EVALUATION PLAN

In order to manage the evaluation process an **Evaluation Plan** should be developed. This should draw together the various aspects from this guide. The evaluation plan should include an outline of the preferred research methods to be used. An evaluation plan should therefore include:

- introduction and background to the project/programme;
- type of evaluation
- level and type of support provided through the project/programme and number of beneficiaries;
- what is to be evaluated? In particular, details of who the evaluation is for and aims, objectives and targets that it will need to report against, and how the evaluation evidence will be used;
- monitoring data and other information/feedback available for the evaluation;
- past evaluation evidence, or evidence from evaluations of similar projects or programmes (internal or external);
- timescales – start date and final deadline;
- management roles/responsibilities;
- whether a survey of other primary research with beneficiaries is required; and
- whether the evaluation is to be conducted internally or by external consultants.

If the evaluation is conducted externally by consultants, the points above can be used to formulate the tender specification. The additional aspect that then needs to be established at this stage is the budget for the evaluation. The consultants will then be required to provide the remaining detail required for the evaluation in response to the brief issued for the evaluation, including:

- research design;
- research questions; and
- methodology (including who is to be consulted and how, sample sizes, etc).

If the evaluation is to be conducted internally the project/programme owner will need to identify all of the aspects set out above.

External consultants can provide an independent viewpoint, specialist expertise and a best practice context. However, it is important that relevant data is also collected internally and provided to the consultants. This will make the evaluation process run more smoothly and avoids duplication of effort.

3.4.2 EVALUATION TERMS OF REFERENCE

If the evaluation is to be undertaken by external consultants it will be necessary to issue **Terms of Reference**. As discussed above, the information in the evaluation plan can be used for the Terms of Reference. It should also include some additional information, including:

- deadline for submission of proposals;
- what information the proposal should provide;
- process for assessing the proposals and short listing; and
- interview process if required (if a date(s) for this have been set they should be included in the brief).

If a date for the inception meeting has been set, this should also be included.

Scottish Enterprise has specified procurement procedures which must be carefully followed when developing the evaluation specification, interviewing, appointing and contracting with consultants. However, a general indication of the content of the Terms of Reference, from the standpoint of evaluation, is noted in Table 3.1 below.

Table 3.1 Terms of Reference Outline

Terms of Reference Outline
Introduction and background to the project or programme and the evaluation (including any relevant regulatory framework)
Details as to who the evaluation is for (main users and stakeholders of the study)
Goals and scope of the evaluation
Evaluative and Research questions
Available knowledge
Research design and methods to be used
Timescales
Indicative Costs/Budget (depending on procurement rules)
Management/roles and responsibilities
Required qualifications of the team
Structure of Proposal
Submission rules and adjudication criteria

3.4.3 EVALUATION PLAN AND TERMS OF REFERENCE CHECKLIST

The following checklist is provided to assist in planning evaluation; a number of the items are covered in greater detail later in this guidance note.

Table 3.2 Evaluation Plan and Terms of Reference Checklist

Key Steps	Check
1. Introduction and background to the project or programme and the evaluation	
This describes the background to the project or programme including content and delivery, target group, how long it has been running and still has to run, management and reporting structure, progress to date, key issues and details of existing documentation that can guide the evaluation.	
2. Details as to who the evaluation is for	
This defines the main audience for the evaluation to guide decisions on aims, research design and methodology.	
3. Goals and scope of the evaluation	
This describes why the evaluation is being undertaken and what is to be evaluated. This will influence the evaluation research questions that are to be posed.	
4. Evaluation Research questions	
This sets out the research questions that the evaluation will need to answer. This will require a detailed understanding of the stage that the project or programme is at and how far the evaluation will need to focus on content and delivery issues as opposed to a scientific assessment of impact.	
5. Evaluation Research design and methodology	
The research design and methods that are chosen will be dependent on the types of questions that are being posed, the budget	

Key Steps	Check
for the evaluation, expertise and the timing of the evaluation.	
In specifying the research design and methodology, it will be important to outline the types of information that need to be collected, methodological approaches and methods of analysis (including the use of statistical packages such as MS Excel or SPSS).	
In considering the research design, it would be useful to consider the following:	
the size of the project/programme or the stage that the project/ programme is at and, thus, the priority/importance of the evaluation;	
whether the evaluation should be primarily qualitative or quantitative and how methods can be triangulated;	
sensible sample sizes and the sampling strategy (particularly important for quantitative evaluation studies);	
Data Protection and ethical issues;	
whether the evaluation is to be undertaken by external consultants or internally and, if internally, how far it is to be a learning experience for staff; and	
the nature of the internal monitoring and evaluation data available.	
It is important to include details of any previous reviews, customer satisfaction surveys, baseline statistics, evaluations, CRM and monitoring data that might be drawn on.	
6. Timescales	
It will be necessary to set a start date for the evaluation in addition to a final deadline. It would also be useful to set deadlines for specific activities including the desk research, data collection, analysis, reporting and dissemination stages.	
7. Costs/budget	
Depending on which department is to fund the evaluation (generally the SEER group and budget), this will need to be agreed internally either by the Appraisal and Evaluation team or the project or programme manager.	
8. Management/roles and responsibilities	
It will be important to decide who is responsible for managing or undertaking the evaluation and what the reporting lines are. Additionally, it will be necessary to clarify who will be responsible for ensuring that findings and recommendations are acted upon and that they feed into the project and programme development and planning processes.	
9. What should the Evaluation Terms of Reference include?	
The Evaluation Terms of Reference will need to include much the same information as the Evaluation Plan but may not require the same amount of detail in some areas. For example, a decision may be made to keep the details of the methodology and budget vague in order to gather the consultants' own suggestions for these areas. However, it is important to stress that the less directive the guidance, the higher the likelihood of receiving proposals that are not comparable and do not meet the needs of the evaluation.	
In drawing up an Evaluation Terms of Reference for external consultants, it will also be necessary to specify deadlines and timescales relating to:	
submissions of proposals;	
assessment of proposals and short-listing;	
interviews;	
selection of successful consultant;	
and the date for the inception meeting if it has been set.	

3.5 HINTS & TIPS

The following general guidance is noted to assist project owners and evaluators in planning for evaluation.

Table 3.3 Hints & Tips for Planning Evaluation

Planning Evaluation	
1	Remember that we evaluate in order to improve projects and programmes – not to undertake evaluations for their own sake. Always ask when planning an evaluation: how will the results improve the lives of Scottish Enterprise customers, the prosperity and well-being of Scotland and the competitiveness of different economic actors. If you cannot find a positive answer to these questions, maybe you should look again at the need for an evaluation or at the very least, at the way it has been designed.
2	Aligning the time cycles of evaluations with the time cycles of programmes and policies is a worthy goal! This is the way to ensure evaluations

Planning Evaluation	
	make their maximum contribution. It is better to deliver an incomplete or imperfect evaluation on time than to achieve a 10% improvement in evaluation quality and miss the window of opportunity, when policy makers and programme managers can use evaluation results.
3	Different stakeholders e.g. policymakers, professionals, managers and customers, have different expectations of evaluation. If a major stakeholder interest is ignored, this is likely to weaken an evaluation, either because it will be poorly designed and/or because its results will lack credibility. Involving policy makers and those responsible for programmes will ensure they take results seriously. Identify your stakeholders, find out what their interests are in an evaluation and involve them!
4	Evaluations must be fully integrated into project/programme planning and management. Programme managers need to think of evaluation as a resource: a source of feedback, a tool for improving performance, an early warning of problems (and solutions) and a way of systematising knowledge. Evaluation is not simply an external imposition. Of course, this truism has implications for evaluators, who need to take on board the concerns of project/programme managers (and their partnerships) and try to take seriously their need for answers to difficult questions.
5	Getting good work from the diverse groups which make up the contemporary evaluations professional community needs bridge building and team building. Bridges need to be built at national, regional and international levels between the different traditions among evaluators – social scientists, economists, policy analysts and management consultants. So hold conferences and support professional exchange to ensure the diffusion of knowledge and know-how. This is one way of building capacity. At a micro-level, the priority is integration and the combination of different skills and competences within evaluation teams.
6	Evaluation is not only about looking back to rate success or failure and allocate blame. It has a contribution to make at every stage in the programme cycle. In particular, evaluation can at the earliest stage, strengthen programmes by helping to unpick intervention logics and reveal weaknesses in programme design – allowing remedial action to be taken early.
7	It is no longer acceptable to gather large quantities of data in the belief that these will eventually provide answers to all evaluation questions. Data dredging is nearly always inefficient. This does not mean that data systems are not essential: they must be put in place at an early stage. However, by being clear about assumptions, by drawing on available theory and being clear about the type of evaluation that is needed, evaluations can be more focused and offer a higher yield for the resources expended.
8	The policy context is an important framework within which evaluations need to be located. Of course, policy changes, or is being constantly restated in different terms and with subtly changing priorities. However, it is always necessary to keep one eye on the policy debates and decisions in order to ensure that evaluations are sensitized to policy priorities. The broader criteria that need to be designed in to evaluations usually derive from the wider policy framework.
9	Although we have argued that all stakeholders are important (see 3 above), the emphasis on socio-economic development gives particular prominence to one important and often neglected group: the intended beneficiaries of the programme interventions. Incorporating the voice of these intended beneficiaries – established Scottish companies, inward investors, new economic entities – in evaluations implies more than asking their opinions. It also implies incorporating their criteria and judgements into an evaluation and accepting that their experience and benefits are the justification for programme interventions.
10	Be pragmatic! We live in an imperfect world where resources are limited, administrators are not always efficient, co-ordination is imperfect, knowledge is patchy and data is often not available. It is nonetheless worth taking small steps, working with what is available and increasing, even marginally, the efficiency and legitimacy of public projects and programmes. Even modest outputs can make a big difference – especially when this is set within a longer-term vision to build capacity and allow for more ambitious evaluations in the future.

4 DETERMINING EVALUATION RESEARCH DESIGN AND METHODS

4.1 EVALUATION RESEARCH DESIGN

It is useful to identify the difference between **research methods** (the set of ways of undertaking research) and **research methodology, or research design** (the set of methods and principles underpinning a specific piece of research).

The overall approach to the research, the broad research design, falls into two general categories: **scientific designs**, which use **quantitative** methods, and **'naturalistic' designs** which use **qualitative** methods.

The design used will depend on the type of project or programme that you are evaluating and the key research questions that you want to answer. It will also depend on the size of the budget that has been allocated to the evaluation. It is also worth noting that much benefit can be gained from a combination of the scientific and naturalistic design.

4.2 SCIENTIFIC/ QUANTITATIVE DESIGNS

Scientific research designs involve the collection of **quantitative data** (i.e. 'metrics', 'measurements' or 'numbers') from larger samples and these designs generally allow more robust generalisations to be made about the population being researched. The main research instrument used in scientific designs, and relevant to Scottish Enterprise evaluation, is the survey questionnaire. The results of which are usually analysed using a software package such as SNAP, SPSS or MS Excel.

Technical Note on Research Design

It is worth noting that there are two main types of scientific designs: experimental and quasi-experimental designs. **Experimental** designs are where subjects are randomly assigned to an experimental or control group so that the net effects of the project or programme can be explored through making comparisons between the experimental and control groups.

However, in evaluation research, it is rarely possible to undertake a true experiment, not just because random assignment is hard to achieve but because there are ethical issues to consider in withholding an intervention from a control group. Such designs usually also require a substantial amount of statistical knowledge and expertise.

Quasi-experimental designs are those in which random assignment of subjects to experiment and control groups has not been achieved. Time series, or longitudinal, designs are quasi-experiments which involve the collection of data on the same population or cohort of individuals, in a comparative way, at regular intervals during a given period. Data are obtained by observing the changes of characteristics or circumstances of those being observed. There is not a requirement for these designs to involve a control or comparison group, the experimental group is enough. Scottish Enterprise evaluations often take this form.

Commonly these are in the form of either **Panel Design** or **Retrospective Panel Design**. In the former, the same group are looked at over a period of time, during which an intervention takes place. This means you need to keep track of people over time. In the context of evaluation this assumes a monitoring and evaluation framework has been put in place at the appraisal stage. This is in practice, not always the case.

In some cases it is not feasible to track the group, or a suitable monitoring exercise has not been established at the outset to enable it. In this case we may ask information at one time only to find out about events in the past e.g. at the end of a business support intervention (retrospective panel design). The obvious disadvantage is reliance on memory and perception of past events.

4.3 NATURALISTIC/ QUALITATIVE RESEARCH DESIGN

Naturalistic inquiry is the study of real life situations as they unfold naturally. It seeks to be unobtrusive and non-controlling and there is openness to whatever emerges with a lack of predetermined constraints on outcomes. Advocates highlight the benefits of **qualitative methods**, which aim to understand the experiences of beneficiaries in-the-round and within in the context of a specific setting.

Examples of these participatory rather than detached methods include interviews, case studies and focus groups. They elicit open-ended responses and tend to result in rich, in-depth understanding of participants' experiences of the project or programme, individual variation, process and quality issues and the effect of context on outcomes.

A positive feature of such approaches is that data can be collected on areas that aren't easily quantifiable, such as changes in perceptions, attitudes and improvements in confidence, motivation and self-esteem, in addition to process issues. Categories and dimensions emerge via the research rather than being pre-determined as in the experimental approach.

This research approach aims at elucidating and understanding the internal dynamics of how a project, programme, organisation or relationship operates which aids replication. Although, due to the small sample sizes, which are usually used when undertaking qualitative research, it is harder to generalise the findings from qualitative research designs to the wider population and questions of bias can arise.

4.4 COMBINING METHODS: TRIANGULATION

All methods and techniques have both strengths and weaknesses. It is generally thought that the best research designs are those that use a skilful mix of scientific and naturalistic approaches using both quantitative and qualitative methods. This use of different approaches is called **Triangulation.**

Triangulation means combining research methods to give a range of perspectives. It is often beneficial when designing an evaluation to incorporate aspects of qualitative and quantitative research designs. For example, as well as survey beneficiaries you may also be interviewing project managers and deliverers.

Technical Note on Triangulation

Triangulation can refer to three main types. Data triangulation involves data being collected at a variety of times, in different locations and from a range of people and organisations. Investigator triangulation involves the use of several rather than just one observer of the same object.

However, **Methodological triangulation** is the most important type. This involves the following:

'Within method' triangulation involves different types of questions within one research instrument e.g. a combination of attitude scales, forced choice questions and open-ended questions within the same survey;

'Between method' triangulation involves the use of a range of research methods.

For example, this might involve combining a large-scale quantitative survey with focus groups and in-depth interviews. This would allow generalisations from a large sample to be made in addition to an in-depth picture of the workings of the project/ programme and the context within which it is operating to be gained.

4.5 RESEARCH DESIGN & METHODS FOR DIFFERENT TYPES OF EVALUATION

The chosen research design and methods will vary by specific focus of the study in hand, but as general guide the table below (Table 4.1) provides examples of the type of approaches relevant to different levels of evaluation.

Table 4.1 Research Methods by Evaluation Type

Evaluation Type	Outline Scope*
Strategic Review. An early intervention assessment sufficiently soon after start date to allow intervention changes if needed e.g.1 year from the intervention start date.	• Assessment of process efficiency (administrative, managerial, technical); • Review of monitoring data; • Qualitative assessment of strategic fit and rationale of project. • May also be repeated at regular intervals for low cost/risk, non-mission critical projects
An **interim evaluation** e.g. 2-3 years into intervention delivery.	• As above, plus assessment of qualitative and quantitative achievements in terms of outputs and immediate benefits against the main intervention objectives; • May include independent review.
A **full impact evaluation** following completion of the intervention, or following a sufficient period after intervention start to capture the significant intervention benefits (e.g. 5 years).	• As above, plus assessment of the additionality of outputs and the wider intervention outcomes and impact; • Qualitative and quantitative assessment of effects of the intervention against the main project themes; • It is recommended that the evaluation includes independent review.

Source: author. Note: *specific requirements will vary by intervention type and phase.

4.6 PLANNING EVALUATION RESEARCH

Figure 4.1, below, provides a flowchart indicating the main stages to go through in undertaking or commissioning evaluation research.

The first stage is generally to make use of available secondary information (often referred to as **'desk-based research'**). A key aim of evaluation is learning from experience. Scottish Enterprise has undertaken a large number of evaluations and therefore has a wealth of experience upon which to draw. These are not only a good source of findings and lessons to inform the future delivery of projects/programmes but also in terms of the process to be adopted in evaluating them. These previous evaluations can be accessed at www.evaluationsonline.org.uk.

Similarly full use should be made of external sources of evidence, e.g. evaluations from other organisations, academic research, official research and reports, from both elsewhere in the UK and abroad.

As noted in Figure 4.1, desk research may be sufficient for some studies, however, generally speaking, most evaluation will entail the creation of new evidence, in other words, **primary research**.

Primary research involves collecting specific information for a particular purpose, as opposed to using information which is readily available or has been previously collected. Sometimes, primary research is required to provide the information that is needed to underpin the overall study and any recommendations or actions arising.

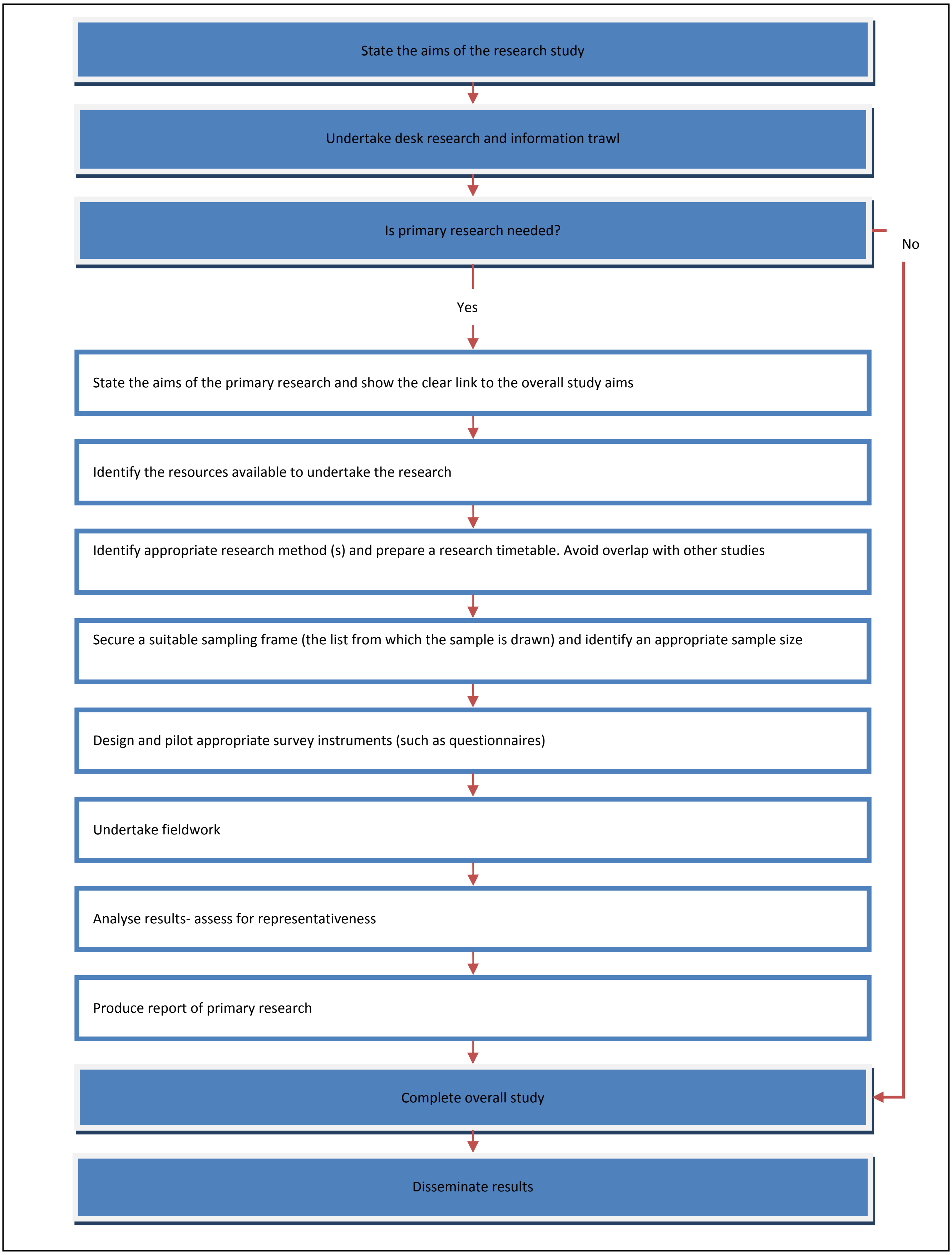
Primary research can be a technical process, but many aspects are straightforward to undertake. The Appraisal & Evaluation team can advise where professional input can be most beneficial in primary research.

The benefits of bespoke primary research are:

- The information gathered can be specific to the topic of research;
- Localised information can be gathered at the appropriate geographical level;
- The information gathered is current (i.e. up-to-date); and
- In addition to quantitative information (numerical data), qualitative information can be gathered (relating to perceptions and opinions)

The main drawback to primary research is cost. Gathering statistically reliable information can provide a drain on scarce resources. It is important that primary research is fully justified and led by a clear need, rather than research for its own sake.

Figure 4.1 Research Flowchart



4.7 EVALUATION RESEARCH PLANNING CHECKLIST

The checklist below provides some useful guidance on thinking through the primary and secondary research requirements of your study. The following sections look at desk research and primary research, in the context of Scottish Enterprise evaluation, in more detail.

Table 4.2 Pre-research Tasks Checklist

Step	Activities	Check
1	• Before thinking about the balance to be struck between desk research and primary research, have a clear notion about the purpose of the proposed study. • A formal statement of your study aims and goals may prove useful at this stage;	
2	• Undertake a trawl of the available information. This should include official information sources, previous local studies, national research and detailed studies from other areas on the same topic.	
3	• Additionally, it should be worthwhile assessing what information may be available from local partner organisations. • A thorough search of the information available will help prevent unnecessary primary research, ensuring that the primary research conducted demonstrates added value;	
4	• If, after the above considerations, primary research is still required, consider what, exactly, you must find out, to contribute to the study aims and goals. • Proper definition of the aims and goals of the primary research (and a clear link to the overall study aims) will prevent a waste of resources;	
5	• Consider what resources you have available to undertake the primary research required. • As stated, partnerships can be useful here in broadening the range of resources available. It is sometimes possible to 'piggyback' some questions on existing planned research, although it is important to balance this against any possible loss of depth or focus in the research. • Ensure that the resources allocated are adequate, and are of an appropriate scale when compared to the aims of the study;	
6	• Think at this stage about the appropriate research 'tools' and methods which should be used in the primary research. • These should be based upon the aims of the study balanced with the resources available;	
7	• If a large-scale study is being considered, check with relevant national and local research organisations to minimise duplication and overlap. • It is important to ensure, as far as possible, that there is no overlap with existing research planned or underway;	
8	• Have a clear timetable identified and allow sufficient time for research piloting and chasing-up responses. • Project planning tools such as GANTT charts or Critical Path Analysis can be useful at this stage.	
9	• Think about dissemination mechanisms and promotion of the research results, bearing in mind the ethical issues, particularly relating to respondent confidentiality. • It is worth remembering that your research may prove useful for reasons other than its intended purpose. • Other researchers may find your primary research and its conclusions useful when conducting their own information trawl as part of their research studies.	

5 DESK RESEARCH

Using secondary information (often referred to as **'desk-based research'**) is a valuable approach in economic analysis. This approach often provides a relatively low cost method by which to investigate particular issues.

However, it is limited by the fact that the researcher must use 'the information that is available' rather than 'the information that is needed'. Nonetheless, it is essential to start any evaluation assignment with a review of secondary data sources including:

- Baseline studies (these will allow you to compare the current position of beneficiaries and the socio-economic context with that at the start of the project or programme);
- Internal management, monitoring, customer satisfaction and evaluation data;
- Project/programme appraisals;
- Other Strategic Reviews, Interim Evaluations, and Full Impact Evaluations; and
- Academic research and studies commissioned by peer organisations.

A key source of evaluation evidence on Scottish Enterprise projects and programmes is Evaluations Online (http://www.evaluationsonline.org.uk). Evaluations Online is a publicly-accessible collection of evaluation reports commissioned or completed by Scottish Enterprise. These reports cover all aspects of Scottish Enterprise's economic development activities and can be searched by key word and theme.

5.1 DESK RESEARCH FOR SCOTTISH ENTERPRISE EVALUATION

In particular, the following areas are ones which are typically useful to examine in Scottish Enterprise evaluation.

5.1.1 DOCUMENTATION REVIEW

The evaluation should begin with a documentation review. This would cover all of the relevant documentation relating to the project e.g. approval papers, funding applications etc. These may help identify the market failure rationale, objectives and targets which the evaluation will assess performance against.

This may also include previous evaluation evidence of the current project/programme or similar projects elsewhere (e.g. Scottish Enterprise evaluations at (see Evaluations Online http://www.evaluationsonline.org.uk/evaluations/eval/advanced.do).

5.1.2 REVIEW STRATEGIC FIT AND CONTRIBUTION

Projects/programmes should be undertaken within the context of strategic frameworks. The relevant strategic frameworks (including internal Scottish Enterprise strategies and external strategies, including the Government Economic Strategy, and the ways in which the project/programme fits with these should have been identified at the appraisal stage.

At the evaluation stage there is a need to review the extent to which the delivery of the intervention has actually contributed to these strategies. Strategic frameworks also develop and change over time. So there is also a need to assess how the project fits and contributes to current strategic frameworks.

5.1.3 MONITORING DATA REVIEW

At the outset of the project, data to be monitored on a regular basis should have been identified. This will also include baseline line data on the beneficiaries so that changes over time can be monitored. The effectiveness of the data being gathered, and its usefulness in providing a reliable indicator of progress towards the achievement of an intervention's objectives should be reviewed as part of the evaluation process.

Data being monitored during the lifetime of an intervention should provide details on:

- the performance of the project; and
- a preferred research framework for the evaluation e.g. who needs to be consulted/surveyed, the volume of beneficiaries to be consulted and what the appropriate sample sizes should be.

Monitoring is a key consideration at the appraisal stage but has implications for the evaluation. So whilst this guidance relates to evaluation it is still worth reiterating here that at a minimum the following needs to be established for monitoring at the appraisal stage:

- what data needs to be gathered to give reliable and consistent measurement against an intervention's objectives;
- who will have responsibility for gathering it;
- when it will be gathered; and
- how it will be gathered and stored.

Where the required monitoring data is not readily available for the project the evaluator needs to:

- highlight this as a shortcoming and identify the reasons for data not being available; and
- take steps, as far as possible, to retrospectively collect and analyse data on the performance of the project (the extent to which this is pursued needs to be balanced against the costs involved).

In most cases this monitoring data would only extend to gross impacts (i.e. not taking into account deadweight, displacement, substitution, leakage or economic multipliers). For an Interim and Full Evaluation this monitoring data would be complemented by data captured through the beneficiary surveys. The beneficiary surveys may then be used to convert from gross to net impacts through the assessment of additionality. For 'Strategic Reviews' the monitoring data will generally be the extent of the quantitative data that will be captured.

5.1.4 WIDER ECONOMIC CONTEXT

Consideration should also be given to the economic environment that currently prevails and that which did so at the appraisal stage of the project/programme. Changes in the wider economic environment may have affected the performance of the project/programme. Taking account of this will set the evaluation within the appropriate context.

5.1.5 LINKS WITH OTHER SCOTTISH ENTERPRISE INTERVENTIONS

The evaluation should consider any links the project or programme has with other Scottish Enterprise interventions. This should include the assessment of any dependencies on, or contributions to, other interventions that may affect performance. Any issues of potential double counting of benefits should be highlighted.

5.2 CHECKLISTS

The two checklists provided below (see Table 5.1 and Table 5.2) cover some of the key issues with desk research. The first is aimed at users who are **conducting** desk research; the second takes a different perspective, and is aimed at those who are **making use of** secondary, desk research, generated by others, e.g. external contractors.

Table 5.1 Checklist for Undertaking Systematic Desk Research

Issue	Check
1. Formulating an Answerable Question	
Does the central question of the review clearly address the following points?	
• The intervention for which evidence is sought	
• The population or sub-groups that the intervention is expected to effect	
• The outcomes that the intervention is expected to achieve	
2. Searching For Relevant Studies	
Have the following steps of a search strategy been planned?	
• The searching of appropriate electronic/internet sources?	
• The searching of appropriate print sources (e.g. journals, textbooks, research reports)?	
• The hand searching of appropriate print sources?	
• Searching of the 'grey' (i.e. unpublished) literature? (e.g. www.evaluationsonline.org.uk)	
3. Critically Appraising Studies Found	
How will the existing literature be sifted for quality and validity?	
• The appropriateness of the questions, populations and outcomes addressed	
• Evidence of selection bias in the primary studies	
• What criteria will be used for including and excluding primary studies	
4. Extracting Data From Included Studies	
Has a strategy been planned for extracting data from the included studies?	
• A data collection form recording how, and why, data were extracted from included studies	
• Information about the characteristic of included studies	
• Details of study characteristics	
• Details of study methods	
• Details of study participants (populations and sub-groups)	
• Details of study interventions	
• Details of study outcomes and findings	
5. Analysing and Presenting The Findings	
• What comparisons should be made (e.g. by interventions studied, participants included, outcomes measured)?	
• What study results are needed for each comparison?	
• What assessments of validity are to be used in the analysis?	
• Is any other data or information needed from authors of studies included in the review?	
• Do the data from different studies need to be transformed for the review's analysis?	
• Is a meta-analysis of findings possible?	
• What are the main findings of the review?	
• What are the main caveats and/or qualifications of the findings of this review?	
6. Interpreting the Findings	
• What is the strength of the evidence from the review?	
• How applicable are the results of the review to 'real life' policy and practice?	
• What does the review say about the costs and benefits of the proposed intervention?	
• What trade-offs are suggested by the review between expected benefits, harm and costs (including opportunity costs)?	
• What mediating factors emerge from the review that might affect the implications for policy and practice in different contexts?	
7. Summarising the Implications for Policy and Practice	
• What are the 'take home' messages for policy making and/or practice?	
• What are the 'take home' messages for future research in this area?	

Table 5.2 Checklist for Using a Systematic Desk-Research

Issue	Check
1. The Question You Want Answered	
Has the review covered the following features?	
• The intervention for which evidence is sought	
• The population or sub-groups that the policy is expected to effect	
• The outcomes that the intervention is expected to achieve	
2. The Adequacy of the Review	
Are there sufficient details in the review about the search strategy used to find relevant research evidence?: i.e.	
• Were appropriate electronic/internet sources searched?	
• Were appropriate print sources (e.g. journals, textbooks, research reports) searched?	
• Was any hand searching of appropriate print sources used?	
• Was an attempt made to identify the ‘grey’ (i.e. unpublished) literature	
3. Critical Appraisal of Evidence	
Were the following tasks undertaken in the review?	
• Was the research evidence that was identified sifted and graded for quality?	
• Were the inclusion and exclusion criteria of primary studies made explicit?	
• Were the appropriate outcome measures included in the review?	
4. Quality of Evidence Presented	
• Is the research evidence presented in the review easy to understand?	
• Is a full evidence table presented?	
• Has the strength of the existing evidence been assessed?	
• Are there any estimates of the likely effect size of the intervention?	
• Are there any details about the contexts in which the intervention is likely to be effective?	
• Is there any information on the likely costs and benefits of the proposed intervention?	
5. Applicability of the Evidence	
• Has the evidence been presented in a way that is helpful to decision making?	
• What are the implications of the review for policy and practice in this area?	
• Are there any mediating factors from the review that need to be taken into account?	
• What are the implications of this review for future research and evaluation in this area?	
6. Peer Review	
• Has this systematic review been peer reviewed by independent analysts with expertise and experience in research synthesis?	
• If not, do you want to get this systematic review peer reviewed?	

6 SURVEY RESEARCH

There is not scope in this guidance note to go into all the research methods that can be used in evaluation. Instead, we have introduced the choices that need to be made both about broad families of methods and highlight key considerations for some of the most common techniques within these families. The first of these is the area of survey research, a commonly used tool for the evaluation.

6.1 FINDING A SAMPLE

A first step in survey research is defining the sample from which to survey. It is rarely the case that data can be gathered for every beneficiary. Therefore, a method for sampling must be established. A sample is obtained by collecting information about only some members of the population or universe of enquiry. Samples can reflect the populations or groups from which they are drawn with varying degrees of accuracy. There are a number of key terms which must be defined prior to undertaking a discussion on sampling approaches (see Table 6.1).

To ensure that a sample is representative of the population or group from which it is gathered, it is crucial that certain types of people in the population are not excluded from the sample. For example, if a household survey was undertaken during the day, it would exclude all those who worked and the sample would be biased. A sampling frame is a list of the members of the population of interest.

There are two broad types of sampling: probability and purposive (or non-probability) (Figure 6.1). Probability samples are generally used in full impact evaluations, where accurate estimates of impacts are required. Generally non-probability samples are sufficient for strategic reviews and for many interim evaluations.

Table 6.1 Sampling Definitions

Term	Definition
Population/Universe	The population/universe refers to the entire group under study.
Sampling Frame	The sampling frame is the list which is taken to represent the population/universe and from which the sample is drawn.
	If the population/universe under study is all companies in an area, then the telephone book or a business directory may be a suitable sampling frame.
Sample	A subset of the population, drawn from the sampling frame, for which information is collected. Inferences about the population at large can then be made from an analysis of this information.
Sample Accuracy	Sample accuracy refers to how confidently we can use the results.
	As an example, it is common to find published work where results are significant 'at the 95% confidence level' - where 95% is the degree of confidence in the results.
Sample Precision	Sample precision relates to the maximum margin of error that can be tolerated in the results produced.
	For some studies, an error of ±5% may be acceptable, in others there may be more stringent requirements.

Figure 6.1 Main Types of Sample

Probability sampling

- Where each person in the population is chosen at random and has an equal, or at least a known, chance of being selected.
- In a non-probability sample, some people have a greater, but unknown, chance than others of selection and, in some cases, the chance of selection is zero.

Non-probability sampling

- Used in situations where probability sampling techniques are either impractical or unnecessary and where a random sample (see below) is not required.
- It may be used when sampling frames are unavailable, or the population is so widely dispersed that random sampling is too inefficient.
- For example, it would be very difficult to obtain a random sample of drug users.
- Purposive sampling may also be used when developing and piloting interview or focus group schedules and is often used in conjunction with probability sampling.

It is important to note that there are many sampling approaches, each of which has its own key strengths and possible drawbacks. Key ones are noted in Figure 6.2 and Figure 6.3.

The choice between these main types depends on:

- Nature of research problem;
- Availability of good sampling frame;
- Resources available;
- Desired level of accuracy; and
- Method of collection.

This guidance can only touch upon such issues (more information on survey and sampling can be found in the source materials listed at the end if this document). Specialist expertise may be required in determining sampling strategies and, in these cases, professional statistical advice sought. As noted, for strategic reviews or interim evaluations, non-probability methods are often most relevant. Project owners and evaluators should be clear on the type of sample to be used in their evaluation, and to what extent it can be used to talk about the study group as a whole.

Figure 6.2 Probability Sampling

Simple Random Sampling

- In simple random sampling, each member of the sampling frame (the list from which all potential samples can be drawn) is consecutively numbered from 1 onwards.
- A list of random numbers (or a random number generator such as that available in standard computer packages such as Microsoft Excel) is used to generate a range of numbers varying randomly between 1 and the size of the sampling frame. This list is then cleaned to remove duplicates, and used to select the appropriate sample size, where the number in the list corresponds to the member in the sampling frame with that corresponding number.
- It should be noted that a random sample is not necessarily representative of the population at large. Although the random approach is adopted to reduce selection bias, it does not by itself mean that any single random sample is necessarily representative of the population at large.

Stratified Sampling (sometimes known as Stratified Random Sampling)

- In stratified sampling, the population under study is divided according to key characteristics. Representatives are then selected at random from within each sub-group (stratum), usually according to the respective proportions in the population at large.
- This approach is often adopted for company surveys, where companies are grouped according to their size in terms of the numbers of employees and their industrial sector. This approach can help reduce bias as the responses are more representative across the strata than a simple random sample may be.

Systematic Random Sampling

- This is similar in approach to simple random sampling.
- However, rather than producing a list of random numbers and then using this list to select corresponding population members from the sampling frame, an alternative selection process is adopted. A random start point is chosen, and a suitable 'step' adopted. From the start point, the next member who is the appropriate 'step' ahead, is chosen and so on, including a wrap-around at the end of the set if necessary.
- As an example, supposed we wished to select a systematic random sample of 20 people from a sampling frame of 100. We would adopt a step of 100/20 = 5. Then we select a random start point using a random number generator (or list of random numbers) - say this is 88. Then we select members 88, 93, 98, 3, 8, 13, 18..., 83. It is important when adopting this method that there is no pattern in the underlying numbering of the sampling frame.

Cluster Sampling

- Cluster sampling involves selecting respondents randomly from chosen clusters.
- This is particularly useful when the population under study is widely dispersed. Specific clusters (e.g. neighbourhoods or community centres) may be selected. Sampling is then performed within these specific clusters, rather than sampling across all available clusters.

Figure 6.3 Non-Probability Sampling

Convenience Sampling

- As the name suggests, this approach is driven by convenience. The sample is selected on the basis of those who are most easily contacted. For example, users of community facilities may be identified and their progression followed (known as a longitudinal study). This approach can prove useful for very detailed and/or longitudinal work but results achieved are not necessarily representative.

Purposive/Judgemental Sampling

- A purposive sampling approach is where members of the population with specific characteristics are identified and invited to participate. If, for example, primary research was conducted into the labour market experience of long-term unemployed young males, then a purposive approach may be a suitable method.

Quota Sampling

- In quota sampling, a precise number of respondents for each sub-group of the population under study is chosen, usually dependent upon the proportion of each sub-group within the overall population. Additional respondents, above the quota limit, in each sub-group are discounted from analysis.
- This approach is similar to stratified sampling, except that rather than weighting responses dependent on the response rate across sub-groups, the quotas adopted ensure that the numbers in each sub-group are proportional to those in the population at large.

Self-Selection Sampling

- This approach relies on responses being received from a self-selecting sample of the population. An example would be customer feedback forms left at restaurants.
- Of the restaurant's customers, only those who wish to fill in a feedback form will respond. This approach suffers from the non representativeness of the responses, although this can sometimes be catered for by weighting responses for known characteristics of the population at large.

6.2 SAMPLE SIZE

It is useful to be aware of some of the main issues affecting the accuracy and precision of probability sample. Required sample size depends on 2 factors:

- Degree of accuracy needed for sample;
- Extent of variation in population in relation to topic being studies.

Table 6.2 shows that we can be 95% confident (or 95 times out of a hundred) that the results in population will be same as in the sample, plus or minus the sampling error. **A figure of 95% accuracy is often taken as a benchmark level of accuracy in socio-economic research**. Results below this level start to undermine the reliability of results.

There are a number of things to note about sample size and accuracy:

a) Sample size
   i. With small samples, a small increase in sample size can give a big increase in accuracy;
   ii. With big samples, increasing sample size does not have the same payoff;
   iii. Rule of thumb: to reduce sampling error by half, need to quadruple sample size;
   iv. Beyond a certain point, increasing sample size not worth it-it is rare to go beyond 2000 (unless accuracy in sub-samples is required).

b) Size of population from which we draw sample largely irrelevant. It is absolute size of sample that is important. **(only if your sample is a big proportion of the total population can you reduce sample size**).

c) Figures in Table 6.2 assume heterogeneous population (e.g. half say 'yes', half say 'no'). If this is not 50/50, you don't need such a big sample- but this is hard to know in advance, so best to assume 50/50- which is a 'worst case'.

d) Of course desired accuracy is not only factor in working out sample size- cost and time are also key factors. The final sample size a compromise between accuracy, cost, and time constraints.

Table 6.2 Sampling sizes required for various sampling errors at 95% confidence level (simple random sampling)

Sampling error %	Sample Size
1.0	10,000
1.5	4,500
2.0	2,500
2.5	1,600
3.0	1,100
3.5	816
4.0	625
4.5	494
5.0	400
5.5	330
6.0	277
6.5	237
7.0	204
7.5	178
8.0	156
8.5	138
9.0	123
9.5	110
10.0	100

Note: table above assumes 50/50 split on the variable being examined. These sample sizes would be smaller for more homogeneous samples (so it's a worst case scenario).

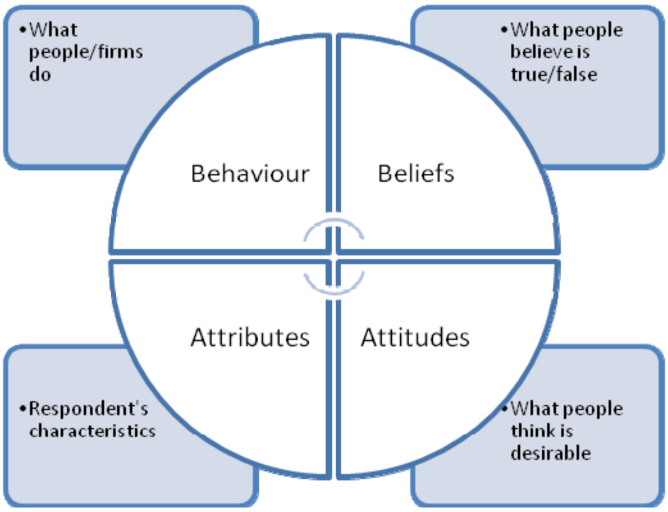
7 CONSTRUCTING QUESTIONNAIRES

Regardless of the type of sampling approach undertaken, an important element of the survey process will be the development of the questionnaire. Careful consideration of survey instruments (i.e. questionnaires) can greatly enhance the effectiveness of primary research. Where primary research has been commissioned, or is being undertaken, it should be standard practice for all survey instruments to be agreed with the steering group, prior to undertaking any primary research.

7.1 QUESTION CONTENT

Consider the content of the question: be clear about what you are interested in. Will the information collected tell you what you are interested in? Knowing the different types of questions can help you be systematic in asking a range of questions. Asking different questions on a topic may reveal interesting things or allow you to combine measures into more powerful composite measures (e.g. combining attitude measures into an index).

Figure 7.1 Question Content



7.2 WORDING QUESTIONS AND GENERAL QUESTIONNAIRE DESIGN

Close attention to the wording of questionnaires is essential. The checklists in Table 7.1 and Table 7.2 provide key issues to be aware of the design and development of questionnaires.

Table 7.1 Question Wording Checklist

Item		Check
Is the language simple?	Avoid jargon and technical terms	
Can the question be shortened?		
Is the question double-barrelled?	Asking more than one question at once e.g. 'How often do you visit your parents?'	
Is the question leading?	Try to avoid implying a preferred answer, e.g. providing a limited range of responses, using loaded words.	
Is the question negative?	Question which use 'not' can be difficult to understand	
Is the respondent likely to have the necessary knowledge?		
Will the words have the same meaning for everyone?		
Is there prestige bias?	For example if an opinion is attached to an important person and someone is then asked to comment on the opinion	
Is the question ambiguous?		
Do you need a direct or indirect question?	Particularly useful for sensitive topics.	

Item		Check
Is the frame of reference for the question sufficiently clear?	For example, do timescales relate to days, weeks, years?	
Does the question artificially create opinions?	Offer a 'don't know' or 'no opinion' options	
Is personal or impersonal wording preferable?	Personal wording on the respondent feels may measure personal attitudes, whereas impersonal questions on how 'people' feel may provide respondent perception's of others attitudes.	
Is the question wording unnecessarily detailed or objectionable?	For example detailed information on age or money. An alternative is to use broad categories.	
Does the question have dangling alternatives?	Providing options before the main subject matter. The subject matter should come before options are listed.	
Is the question likely to produce a response set?	For example, where respondents may think it socially desirable to provide certain answers, it is important to make them comfortable about providing any answer.	

Table 7.2 Questionnaire Design Checklist

Design Issues	Check
• The purpose of the research should be borne in mind when designing survey instruments. • Weed out irrelevant questions. • Resist the temptation to conflate issues ("the kitchen sink approach") and focus on the information required. • Choose carefully where to strike the balance between depth of information compared to breadth of information;	
• Remember the importance of presentation for written surveys. • Provide concise instructions, clearly separate answer boxes/spaces from questions and provide examples in the form where relevant. • For telephone surveys, always have a standard introduction stating the purpose of the research, who has commissioned it, who is undertaking it on their behalf (if relevant) and state the confidential nature of responses as necessary;	
• Recognise the characteristics of your target population. • Technical phrases and jargon may not be understood by some respondents. • Some sub-groups of the population may be better reached by one survey method than another;	
• Look at official questionnaires (the Labour Force Survey, the Scottish Household Survey) for ideas and examples of how it's done in practice. • This is also useful for 'lifting' questions, enabling your primary research results to be compared to national results;	
• The survey method chosen should affect the design of survey instruments. • Telephone surveys, for instance, should not last longer than 10-15 minutes, as a general rule. • Additionally, when using this approach, the respondent should be able to answer any questions then and there, without recourse to 'looking the information up'. • For paper-based questionnaires, careful layout of questions can generate maximum value and reduce the length of the overall questionnaire;	
• Balance the length of the survey with the information required. • Convoluted or overlong survey instruments may reduce the overall response rate;	
• Avoid loaded questions, leading questions with preferred or anticipated responses and rephrase ambiguous questions. • Often, piloting a survey instrument amongst a small sub-group of the population to be targeted pays enormous dividends in the quality of the eventual questionnaire. This, in turn, leads to improved results;	
• Choose a careful balance between open-ended questions (where a range of answers can be given) or closed questions (normally Yes/No answers). • For the former, choose carefully whether to pre-code responses prior to survey (providing a list of options) or to code responses post survey;	
• Pre-coding is quicker, although post-survey coding can now be automated to some extent with appropriate software and skilled interviewers- a balance needs to be struck between practical steps such as pre-coding, and making sure answers are not forced. • Note that pre-coding responses can also include a residual category (such as "Other - please state") which can be coded post-survey. • Again, piloting a questionnaire can assist with identifying suitable pre-coded responses;	

Design Issues	Check
• Consider choice of scaled responses (a Likert scale). These can be a valuable tool, but require careful consideration of the type and meaning of information obtained	

7.3 PILOT TESTING

Questionnaires must be tested rigorously. Pilot testing is probably one of the most underestimated aspects of questionnaire design. It generally pays off not to rush this stage and to do some thorough ground work. There are three phases to pilot testing as shown in Figure 7.2. Table 7.3 provides a checklist for undertaking a through pilot of a questionnaire.

Figure 7.2 Piloting Phases

Phase 1 Question Development
- Testing individual questions

Phase 2 Questionnaire Development
- Administration and evaluation of whole questionnaire
- Often an 'undeclared' simulation

Phase 3 Polishing Pilot Test
- Revisions from phase 2, e.g. shortening, reordering, filerting set-up, and general layout

Table 7.3 Pilot Testing Checklist

Item			Check
Pilot Testing Items			
	Answer variation	Questions producing similar answers and low variation will be hard to analyse and produce some categories with a low number of cases	
	Meaning	Check to ensure that respondents understand the intended meaning of the question and that you understand the respondent's answer	
	Redundancy	If two questions measure virtually the same thing, only one is needed in the final questionnaire	
	Scalability	If a set of questions is designed to form a scale or index, check to ensure that they do.	
	Non-response	Refusal to answer by a large number of people will affect analysis and reduce sample size	
	Acquiescent response set	This occurs where respondents agree/disagree to a statement regardless of content. Where this is suspected replace the 'agree-disagree' format.	
Pilot Testing Questionnaires			
	Flow	Ensuring the questionnaire progresses smoothly and logically from one section to another	
	Question skips	Where filter questions are used, it is important to ensure that the skip patterns work correctly	
	Timing	Time each section of the questionnaire to allow appropriate editing as well as estimating length for budgeting purposes and informing respondent of questionnaire length	
	Respondent interest and attention	Where there is obvious boredom amongst respondents, questionnaire length, and question variety may need to be considered	
Pilot Testing Approach			
	Deciding who to pre-test	The closer the match between the pilot sample and the final sample the better.	
	Number of pre-tests	Too few responses mean that issues such as non-response, variation, response sets etc remain undetected. Ideally 75-100 respondents provide a useful pilot test. However, this may be difficult where the population is small.	
	Deciding who should interview	A selection of interviewers that represent the range of experience of those who will finally administer the questionnaire will provide the most realistic simulation.	
	Coding responses	Try coding phase 2 questions (especially open and 'other' responses to closed questions). Difficulties in coding may highlight problems with question wording or interviewer understanding.	
	Interviewer debriefing	It is essential to learn from interviewer experience with the questionnaire as well as from	

respondents' answers to questions.	

7.4 QUESTIONNAIRE ADMINISTRATION

A further issue to consider is the most appropriate means to deliver the questionnaire, which will also influence the design process. It is important to weigh up the advantages and disadvantages of main delivery options, as they all have strengths and weaknesses in relation to:

- Response rates;
- Ability to obtain a representative sample;
- Effects on questionnaire design;
- Quality of answer; and
- Ease of implementing the survey.

A summary of three common approaches is provided in Table 7.4.

Table 7.4 Advantages and Disadvantages of Mail, Personal and Telephone Methods of Administering Questionnaires

	Face to face	Telephone	Mail
Response rates			
General samples	Good	Good	Satisfactory
Specialised samples	Good	Good	Satisfactory
Representative samples			
Avoidance of refusal bias	Good	Good	Poor
Control over who completes questionnaire	Good	Satisfactory	Good
Gaining access to the selected person	Satisfactory	Good	Good
Locating the selected person	Satisfactory	Good	Good
Effects on questionnaire design			
Ability to handle:			
Long questionnaires	Good	Satisfactory	Satisfactory
Complex questionnaire	Good	Poor	Satisfactory
Boring questions	Good	Satisfactory	Poor
Item non-response	Good	Good	Satisfactory
Filter questions	Good	Good	Satisfactory
Question sequence control	Good	Good	Poor
Open ended questions	Good	Good	Poor
Quality of answer			
Minimise social desirability responses	Poor	Satisfactory	Good
Ability to avoid distortion due to:			
Interviewer characteristics	Poor	Satisfactory	Good
Interviewer's opinions	Satisfactory	Satisfactory	Good
Influence of other people	Satisfactory	Good	Poor
Allows opportunities to consult	Satisfactory	Poor	Good
Avoids subversion	Poor	Satisfactory	Good
Implementing the survey			
Ease of finding suitable staff	Poor	Good	Good
Speed	Poor	Good	Satisfactory
cost	Poor	Satisfactory	Good

7.4.1 SELF-ADMINISTERED TECHNIQUES

Self-administered approaches to primary research (where the respondent completes the appropriate survey instrument unaided) are a convenient way of collecting information from a range of respondents at relatively low cost. However, a burden is being placed upon respondents to take the time and trouble to complete any such questionnaire and this should be uppermost in the researchers mind if advocating this technique. It is therefore important that steps are taken to encourage responses as far as possible. Appropriate factors to bear in mind when considering such an approach are indicated in Figure 7.3.

Figure 7.3 Self-Administered Techniques: Considerations

Who should fill it in?

- Proper targeting, such as to a named individual, will identify who is supposed to supply the required information.
- Failing this, adequate guidance, presented in a concise, clear and unambiguous manner should allow appropriate respondents to identify themselves.

What are the benefits to the respondent? What would make them want to fill it in?

- Promotion of the importance of the survey, including the use to which the information gathered is being put, is a good starting point.
- There should be no cost to the respondent in monetary terms in completing the questionnaire. Pre-paid or Freepost envelopes should be made available for responses.
- Similarly, an incentive to respondents may encourage some respondents to participate.

How long will it take each respondent? How hard will it be to complete the questionnaire?

- Considered design of survey instruments will minimise the burden on respondents. This includes components of design and format, as well as of length.
- A co-ordinated approach to primary research in local areas may also reduce the burden on respondents.

Feedback/Actions

- What actually happened as a result of the primary research?
- If this can be successfully conveyed, then those involved in the process can see the benefits of their participation.

A main type of self-administered questionnaire is a postal survey. Advantages and disadvantages of this approach are noted in Figure 7.4.

Figure 7.4 Postal Survey

Postal Survey - Advantages

- Useful for large scale work across wider geographic areas
- Processing of responses can be automated
- Respondent can complete at their leisure
- Permits 'checking up' of facts by respondent in preparing answers
- Responses can be followed up for even more detailed discussions by telephone or e-mail
- Relatively cheap for the quantity of information obtained
- Consistency across survey instruments/data collection

Postal Survey - Disadvantages

- Excludes those with literacy problems
- Responses may be less full than by other methods (no probing) and generally taken on trust
- Response rates deteriorate with conflation of issues (or lack of research focus) and questionnaire length
- Administrative burden on mailing and data input (if not automated)
- Self-selection (misses the apathetic) leads to possible non-response bias

7.4.2 INTERVIEWER ADMINISTERED TECHNIQUES

Some of the main types of interviewer survey are noted in the table below along with character, strengths and weaknesses. As noted previously, it is important to emphasise, that in Scottish Enterprise evaluation, estimation of

economic impact often forms a **central focus**, which often lends itself to the latter two, more structured, systematic approaches.

Table 7.5 Strengths and Weaknesses of Interview Approaches

Type of Interview	Characteristics	Strengths	Weaknesses
Informal Conversational Interview	Questions emerge from the immediate context and are asked in the natural course of things; there are no predetermined questions, topics, or wordings.	Increases the salience and relevance of questions; interviews are built on and emerge from observations; the interview can be matched to individuals and circumstances.	Different information collected from different people with different questions. Less systematic and comprehensive if certain questions do not arise "naturally". Data organization and analysis can be quite difficult.
Interview Guide Approach	Topics and issues to be covered are specified in advance, in outline form; interviewer decides sequence and wording of questions in the course of the interview.	The outline increases the comprehensiveness of the data and makes data collection somewhat systematic for each respondent. Logical gaps in data can be anticipated and closed. Interviews remain fairly conversational and situational.	Important and salient topics may be inadvertently omitted. Interviewer flexibility in sequencing and wording questions can result in substantially different responses from different perspectives, thus reducing the comparability of responses.
Standardized Open-Ended Interview	The exact wording and sequence of questions are determined in advance. All interviewees are asked the same questions in the same order. Questions are worded in a completely open-ended format.	Respondents answer the same questions, thus increasing comparability of responses; data are complete for each person on the topics addressed in the interview. Reduces interviewer effects and bias when several interviewers are used. Permits evaluation users to see and review the instrumentation used in the evaluation. Facilitates organization and analysis of the data.	Little flexibility in relating the interview to particular individuals and circumstances; standardized wording of questions may constrain and limit naturalness and relevance of questions and answers.
Closed-Field Response Interview	Questions and response categories are determined in advance. Respondent chooses from among these fixed responses.	Data analysis is simple; responses can be directly compared and easily aggregated; many questions can be asked in a short time.	Respondents must fit their experiences and feelings into the researcher's categories; may be perceived as impersonal, irrelevant, and mechanistic. Can distort what respondents really mean or have experienced by so completely limiting their response choices.

7.4.3 FACE TO FACE / IN-DEPTH APPROACHES

The section above introduced the 'Interview Guide Approach'. This section provides more detail on that approach.

Face-face or in-depth interviews or consultations aim to collect rich and in-depth data from smaller numbers of people. Their aim is often to uncover information that may add to that collected through surveys such as individual views and perceptions.

Interviews can be undertaken either face-to-face or via the telephone and require that an interview schedule or interview guide is prepared in advance. An interview schedule is more structured and contains a set of questions in a predetermined order adhered to in each interview. This allows comparison of responses. An interview guide lists areas to be covered but leaves the exact wording of questions to the interviewer.

The advantage of interviews is that more flexibility is allowed and skilled interviewers can extract a lot of rich and detailed information. However, interviews, particularly face-to-face interviews, can be expensive and the findings can be time consuming to analyse. Additionally, interviewer bias can creep in and findings can be difficult to generalise.

Face-to-face approaches are useful in gathering information in sufficient depth to get to the heart of many issues. Although time consuming, and difficult to undertake over a wider geographical area, they can collect information on a semi-structured basis, allowing the respondent/interviewee to shape the emphasis on discussions. It is important that some loose structure is developed prior to interview, to prevent the interview degenerating into a general discussion or informal chat.

In conducting face-to-face interviews it is common, when contacting the interviewee to arrange an appointment, to outline the nature and scope of the interview. The appointment, the nature of the study and any preparatory work expected of the interviewee should be confirmed in writing.

This is also a good opportunity to identify any materials you hope to collect from the interviewee at the interview.

Table 7.6 Face-Face Interviews

Face-to-Face Interviews - Advantages	Face-to-Face Interviews - Disadvantages
•Generally high response rate, useful for smaller samples •Greater depth of questioning – interviewer can probe on specific issues and collect perceptions and viewpoints •Versatility – can cover a range of issues and topics •Interviewee can shape discussions •Interview can be followed up for even more detailed discussions by telephone or e-mail •Skilled interviewers can develop rapport with interviewee •Removes literacy and numeracy barriers	•High administrative costs per response achieved •Physical limit to area covered in 1 day •Specialist interviewer skill required – training or expert help sometimes needed •Some respondents prefer anonymous methods such as postal surveys

7.4.4 TELEPHONE APPROACHES

Telephone approaches to primary research have proven increasingly popular in recent years. However, their widespread use in direct marketing ('cold calling') has adversely affected response rates in general.

Nonetheless, telephone surveys allow a much wider geographical area to be targeted in any single day and are cheaper than face-to-face interviews. Use of computer aided telephone interview (CATI) technology allows responses to be input directly into a computerised format suitable for analysis and this can speed up the entire research process.

Conversely, the telephone interview is difficult to sustain for more than 15-20 minutes, respondents are often difficult to pin down to a specific time and it is therefore difficult to schedule the interview.

Additionally, responses must be capable of being given then and there, and although prior warning of the survey to named individuals can increase responses, the scheduling difficulties make it hard for some interviewees to prepare adequately. In some surveys, such as those involving larger companies, it is also the case that a single respondent may not have the overriding view to enable him/her to answer all questions from a position of knowledge.

It is sensible to have a named contact for telephone surveys and to inform them beforehand of the nature and scope of the survey, together with an indication of when you are likely to be calling. It is also important to give some thought to the best person within an organisation for answering questions on the main topic of interest.

Table 7.7 Telephone Interviews

Telephone Interviews - Advantages	Telephone Interviews - Disadvantages
•Cheaper to undertake than face-to-face interviews •Reasonable response rates, although 'cold calling' marketing has reduced the effectiveness of the method in recent years •Computer assisted telephone interviewing can speed up process, including coding and analysis of responses •Can be useful as a follow-up approach to face-to-face interviews, probing specific points •Can collect information at a reasonable level of depth, provided the research is firmly focused	•Relatively impersonal, difficult to generate the same level of rapport that can be achieved through face-to-face interviews •Can be difficult to schedule appropriate times for interviews •Common usage (especially cold calling) has led to respondent fatigue •May introduce bias (those without telephones are excluded from the process, as are ex-directory numbers) •Some respondents prefer anonymous methods such as postal surveys •Requires shorter questions in general •Limited duration for phone call (maximum 15 minutes in practice)

8 OTHER EVALUATION RESEARCH METHODS

8.1 FOCUS GROUPS

Focus Groups have proven a popular method by which to gather primary research into specific issues or by which to gather the views of sub-groups of the population. Focus Groups bring individuals together to discuss issues in an open format. Excellent moderator skills are required to get the best from this approach, and adequate training or specialist expertise may be required to gain maximum benefit from this method.

The main benefits of the Focus Group approach are that it allows participants to direct discussions into their main concerns and issues, and it can collect very detailed and complex information, particularly information of a qualitative nature. Focus Groups are often useful in both forming questionnaires for a larger survey, or in exploring results from such a survey in more detail.

A focus group should, ideally, include between 8 and 12 people who are brought together to engage in a guided discussion about a topic for up to three hours. However, focus groups with young people tend not to last more than an hour and a half.

The focus group method capitalises on the interaction between participants, which enhances and consolidates the data collected. Subjects are selected on the basis of relevance to the topic under study and usually participants do not represent any meaningful population. Groups should be led by an experienced facilitator using a topic guide. Focus groups are useful for gathering sensitive data from disadvantaged and vulnerable groups such as the long-term unemployed and disaffected and disengaged young people. They require careful and unbiased analysis and reporting and can result in rich data and quotations. Focus Group research can also be completed in a relatively short timescale from start to finish.

One of the main drawbacks with Focus Group research is that, even when targeted at specific sub-groups of the population, participants may not be representative of that sub-group at large. Identifying suitable participants is a major element of successful research in this instance. Results from Focus Group research should not be taken as statistically reliable. Rather they are a good indication of the factors underpinning specific issues.

Table 8.1 Focus Groups

Focus Groups -Advantages

- Can be undertaken quickly and results produced in a short timescale
- Provides detailed insights into particular issues, from the perspective of Focus Group participants
- Versatility – discussions can be shaped by participants
- Can be targeted at particular sub-groups of the population
- Can collect information in greater depth - enables an interactive approach to identifying qualitative information such as perceptions and feelings and the reasons underpinning these views

Focus Groups - Disadvantages

- Relatively high administrative costs per response
- Results are not statistically reliable for the population as a whole
- Danger of interviewer/moderator bias. Requires expert assistance or appropriate staff training to deliver effectively
- Some people prefer anonymous methods - bias in participant structure.
- There is often a degree of self-selection amongst participants (i.e. from those invited, can only obtain information from those who choose to attend)
- Invited participants for any sub-group may be significantly different from that sub group in general

8.2 CASE STUDIES

Case studies are often used to gain an in-depth and comprehensive understanding of an issue. The use of quantitative data and quantitative research methodologies' provides a measure of the impact of an intervention. In broader policy context it is important to gain an understanding of why, for whom and under what circumstances interventions achieve their policy objectives. Addressing these evaluation issues requires a more qualitative, participatory research method. Case studies, similar to focus groups, are therefore often used to go beyond the 'numbers'.

It is important to distinguish between Case Studies that are written for 'marketing' purposes and the use of Case Studies as an independent and objective research method. Case Studies can be used as a promotional tool to demonstrate good performance. Used in this way Case Studies offer limited value within a research programme. If meaningful insights are to be gained from Case Studies the approach adopted should be rigorous and stand up to scrutiny.

Within any research design consideration should be given to the use of a number of complementary research methods. This can help test the validity of the evidence gathered. Case Studies are frequently used in this context. They can complement quantitative data by exploring in greater detail issues that have been brought to the fore within the evidence already assembled. Case Studies can therefore build on the findings of large-scale quantitative surveys to gain a better understanding of policy interventions. A key limitation of case studies is the extent to which future policy should be influenced by a single case or a small group of cases. The use of topic guides that give case studies a consistent structure enables general themes to be explored and policy issues to emerge.

Figure 8.1 Case Studies

Case Study- Advantages

- Encoruages a detailed exploration of why and for whom interventions are succeeding or falling short of expectations
- Participatroy approach can hep identify unintended consequences of an intervention
- Enables the exploration of issues to be more flexible and not be constrained by a survey
- Is complementary to other resarch methods

Case Study- Disadvatanges

- Can be relatively expensive and a time-consuming method
- Diffculties in interpreting general policy conclusions from specific cases
- Results are not statistically reliable for population as a whole
- Can be used superficially to promote good performance
- Selction bias can be an issue, care is needed when selecting participating beneficiaries

8.3 PRIMARY RESEARCH CHECKLIST

A checklist has been developed which should enable a fuller assessment of primary research. Note that not all primary research will satisfy all the points on the checklist. The checklist is useful both in designing primary research and in assessing primary research from third parties (e.g. external contractors).

Table 8.2 Primary Research Checklist

Issues	Check
NEED AND AIMS	
Is it clear which organisation(s) or individuals commissioned the research?	
Is there any evidence of a partnership approach?	
Has the individual or organisation undertaking the primary research been identified?	
Has a formal statement of the research aims and objectives been provided?	
Is there evidence of a comprehensive information trawl?	
Have the information gaps been identified?	
Has reference been made to previous studies?	
Has the need for the primary research been clearly established?	
Is the time period over which the primary research was undertaken clearly stated?	
METHOD	
Is it clear which primary research method(s) were adopted?	
Are the research methods chosen suitable?	
Have the limitations of these methods been acknowledged?	
SAMPLING ISSUES	
Is it clear what sampling frame was used?	
Is the sampling frame suitable?	
Are any problems with the sampling frame acknowledged?	
Were any such problems adequately addressed?	
What sample size was chosen?	
SURVEY INSTRUMENTS (Questionnaires etc.)	
Are copies of the survey instruments (questionnaires etc.) included in the report, or made readily accessible?	
Is the survey instrument well designed?	
Does it match the stated aims of the primary research and are the questions appropriate for the method?	

Issues	Check
If questions from national or comparable surveys are included for subsequent analysis, is the phrasing exactly the same?	
DATA COLLECTION	
Were respondents encouraged to participate?	
Could the participation incentive affect the results?	
Were adequate procedures in place to ensure that the data collected was accurate, and which minimise sampling and non-sampling error?	
Were any ethical standards adopted?	
Does the research adhere to the Market Research Society Code of Conduct?	
REPORTING RESULTS	
What response rate was achieved? Is this adequate?	
Is the level of accuracy to be accorded results clearly stated?	
Have the responses been weighted for the final results?	
Are these weightings clearly identified?	
If comparisons are made with results from other studies, is this linked back to specific elements in the survey instrument?	
Is the report in plain English?	
Are both percentages and base numbers included in the report of findings? Using percentages only is often a means to disguise low response rates (e.g. 6 responses from 10 is a 60% response rate)	
Is there a clear distinction between factual results, interpretation and recommendations from the primary research?	
DISSEMINATION	
Have the results of the primary research been disseminated through suitable media, bearing in mind the target audience?	
Are they readily accessible?	
Are they in a suitable format?	
Can you get further results?	
Is there a contact for discussions and comments?	
Are such comments welcomed?	

8.4 HINTS & TIPS

The following general guidelines for planning evaluation research are noted.

Table 8.3 Hints & Tips for Selecting Methods and Techniques

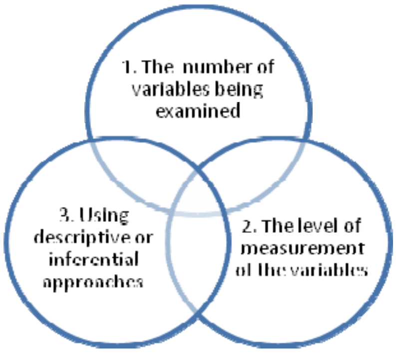
Selecting Methods and Techniques	
1	Choosing methods and techniques follows directly from the kind of questions one wants to ask and these questions are part of an extensive design exercise that includes consulting stakeholders and assessing programmes characteristics. Choosing methods and techniques first and trying to make them fit with questions for which they have not been specifically chosen will always create problems. The techniques chosen need to reflect the purpose and focus of the evaluation.
2	Most techniques have strengths and weaknesses, these need to be recognised and where possible different techniques need to be applied together to strengthen the analysis and make the evaluation results and conclusions more reliable.
3	Because of the distinctive character of socio-economic development: bottom-up, using different combinations of interventions and tailored to territorial and sectoral needs – it is difficult to measure and compare outcomes across socio-economic development programme settings. This is doesn't mean that measurement, quantification and statistics are not relevant. They can be powerful tools when comparisons are at the level of the particular development programme and do not attempt to compare non comparable settings.
4	Qualitative methods and techniques are well suited to socio-economic development because of the subtlety and holistic nature of what is being attempted and because of the differences in contexts which need to be described in qualitative ways. The participatory nature of local development – building on the potential and ambitions of local stakeholders and citizens is especially suitable for both qualitative methods and participatory methods.
5	Thematic priorities which are very common in European programmes pose real difficulties for evaluators. Because policy makers want to understand how far their policies are successful as a whole, there is often pressure to aggregate results and find a common way of describing or even measuring what is happening. This often cannot be done. Sometimes on qualitative descriptions will work. Take care not to add up apple and pears.
6	There is often a tension between choosing evaluators who know a lot about a particular policy area and those whose evaluation skills are more generic. Ideally in an evaluation team one tries to span both of these sets of knowledge and experience. Commissioners need to be aware of the dangers of contracting evaluators who have lived most of their professional lives working in one specialised area and using a limited set of methods and techniques. This is another argument for looking at the balance of skills in the evaluation team.
7	It is important to distinguish between methods and techniques for gathering data, for analysing data and for informing evaluative judgments. This distinction is not always made partly because those who undertake evaluations may be more preoccupied with one stage of the process rather than another. As in all things there needs to be a balance. It is no good investing in sophisticated methods to gather data but to be relatively simplistic in the way data is analysed.
8	Data is never pure or naturally occurring, it needs to be produced. Because of this evaluators need to know where their data comes and what decisions have made in the course of its production. At the end of the day the strength of the arguments and conclusions that can be drawn depends on the strength and characteristics of the data that is being used.
9	One important quality aspect in evaluating data follows from the way data has been accessed and how access has been negotiated. In socio-economic development programmes in particular there are a host of problems to be resolved. Different partners have to be willing

Selecting Methods and Techniques	
	to share information, excluded groups often distrust evaluation as one further example of official behaviour – and need to be brought on board, and all stakeholders need to be convinced that they are going to get something out of an evaluation before they give access with any enthusiasm to any information that they hold. Investing in these kinds of negotiation processes will make a difference to quality and evaluation as a whole.
10	The quantitative/qualitative divide is overstated. Data is often more of a continuum, beginning life as qualitative and once analysed becoming quantitative. Weaker forms of quantitative data (e.g. categorisations or ranking) are close to qualitative data. What is needed when evaluating socio-economic development programmes is qualitative data able to capture subtleties, people's experience and judgements and quantitative data to provide overviews, for example to aggregate results of an intervention and a comparative perspective.
11	Well conceived indicators and monitoring systems are a powerful adjunct to evaluation. Very often evaluators depend on monitoring systems which are indicator based. If these are not put in place early in the programme design cycle it may be too late to create such monitoring systems later on.
12	Over elaborate indicator systems may be counter productive. Whilst there is a temptation in multi-sectoral and multi-objective programmes to measure everything, this should be resisted. This can be costly and the results difficult to use.
13	Indicators are often used for a management and accountability purposes. It can be difficult to reuse indicators that have been exclusively developed for such purposes as part of an evaluation. There can be too much pressure to shape information in positive ways or at the very least make sure that bad news does not come through. On the other hand within a well-developed performance management culture these kinds of indicators can help improve programme content as well as their management. This would appear to be the case in some countries where the performance reserve has been positively incorporated into programmes within the European Structural Funds.

9 ANALYSIS OF SURVEY DATA

Once data have been collected they have to be analysed. This section outlines the broad factors affecting how data can be analysed and what issues you should be aware of for designing/using surveys and survey results. The three main factors to be aware of are noted in Figure 9.1.

Figure 9.1 Factors Affecting Analysis

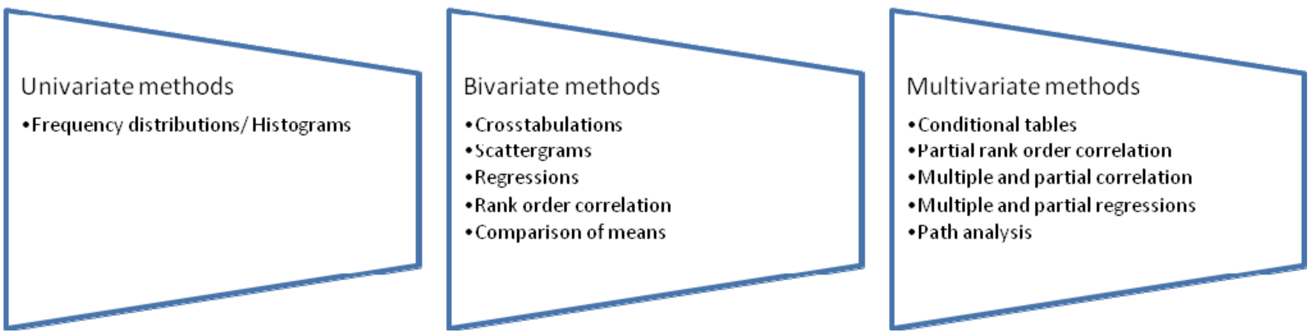


9.1 THE NUMBER OF VARIABLES

How we analyse depends on what we want to know. If we want to look at one variable at a time, this requires a 'univariate' method of analysis. Examining two variables simultaneously represents bivariate analysis. Interest in relations between three or more variables requires 'multivariate' analysis. In practice we develop and refine our research in the process of analysis, so we move between univariate, bivariate and multivariate techniques, as illustrated in Figure 9.2. Some examples of statistical tools are given in Figure 9.3. **Not all of the techniques will be relevant for every evaluation. Simple univariate and bivariate techniques are normally all that is required for Strategic Reviews or Interim Evaluations.**

Figure 9.2 Progressing between methods of analysis

Figure 9.3 Some Methods of Analysis for Survey Research



9.2 LEVELS OF MEASUREMENT

Having decided on the number of variables, we then have to decide the particular technique relevant to these broad categories. A key factor in this choice is the **'level of measurement'** of the variable being used (Figure 9.4).

There are 3 main levels of measurement:

- Nominal/Categorical (an example of Nominal is Yes/No; and example of Categorical is 'red/green/blue');
- Ordinal (which means it is meaningful to rank categories, for example 'agree strongly, agree, neither agree/disagree, disagree, disagree strongly')
- Interval / ratio (which means we can rank and can quantify precisely the difference between categories).

Table 9.1 assists in the identification of levels of measurement for data. Identifying the level of measurement is important as it affects what we can and can't do with the information from our survey (see Table 9.2). **As noted, these issues are generally less relevant for strategic reviews, or interim evaluations, but are important for full impact evaluations**.

Figure 9.4 Main Levels of Measurement

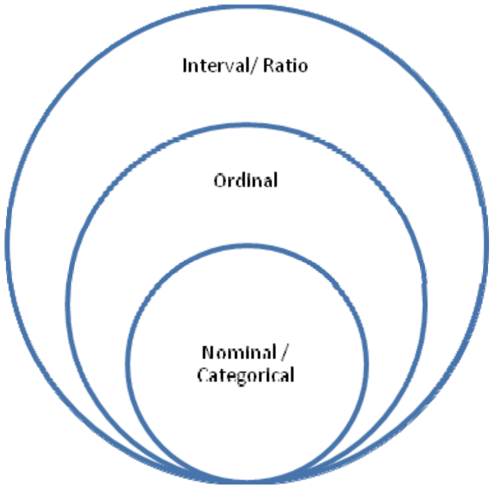


Table 9.1 Working out Levels of Measurement

	Level of Measurement		
	Nominal	Ordinal	Interval
Are there difference categories?	Yes	Yes	Yes
Can I rank the categories?	No	Yes	Yes
Can I specify the differences between categories numerically?	No	No	Yes

Table 9.2 Issues in Selecting Levels of Measurement

Issue	
1	A wider range of methods of analysis is appropriate as the level of measurement of variables increases
2	More powerful and sophisticated techniques of analysis are appropriate for interval level variables
3	Higher levels of measurement provide more information- but is this really necessary?
4	Questions which require a lot of precision and detail can be unreliable since people often do not have accurate, detailed information
5	People may be reluctant to provide precise information but may provide it in more general terms (e.g. income bracket)
6	Numerical data collected in grouped form (e.g. age, income)can be converted to interval data if we make particular assumptions (e.g. normal distribution)
7	If variables are measured at an interval level it is simple to reduce it to ordinal or nominal scale. With the above exception, data collected at low levels of measurement cannot be converted to higher levels.

9.3 DESCRIPTIVE AND INFERENTIAL ANALYSIS

The third main issue to consider in analysis of survey results is type of statistical technique to use: descriptive or inferential (see Figure 9.5), i.e. whether we want to simply describe the patterns in the data, or to make generalisations from the sample results to a wider group of beneficiaries? The key point to note is the progression from univariate to multivariate analysis; and a progression from descriptive to inferential statistics, all placing increasing demands on your data and survey approach: an issue that is something that should be considered at outset, during the research design phase. Ideally, full impact evaluations would seek to gather data of sufficient quality to permit statistical techniques that allow greater insights, i.e. multivariate, inferential techniques.

The flowchart in

Figure 9.6 summaries the methods of analysis appropriate to a given research question.

In Scottish Enterprise evaluations, simple descriptive statistics (with low demands on the level of measurement) are generally sufficient for Strategic Reviews and for Interim Evaluations. However, Full Impact Evaluations will often require greater use, in addition, of inferential statistics (with higher demands on levels of measurement). This places much more emphasis on careful planning of surveys and questionnaires to ensure the right data are collected.

Figure 9.5 Basic Statistic Types

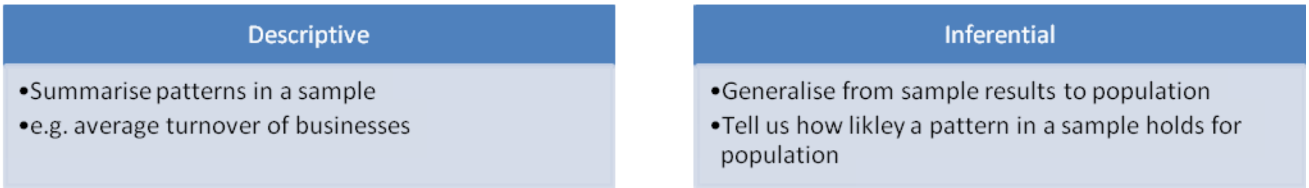
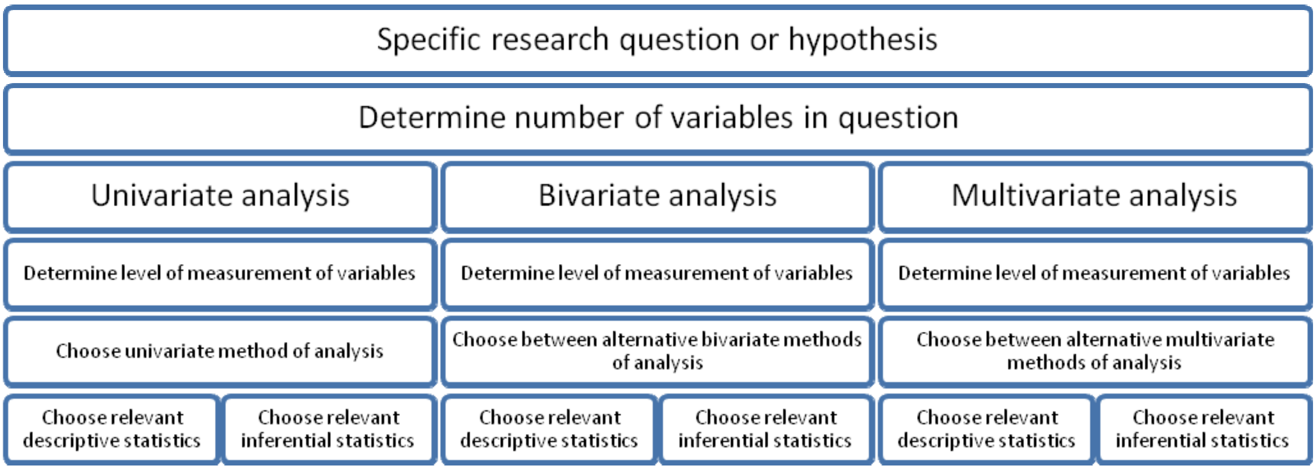


Figure 9.6 Flowchart for Selecting Analysis and Statistical Techniques



10 DATA PROTECTION & ETHICAL ISSUES

10.1 DATA PROTECTION

Evaluation undertaken for Scottish Enterprise must honour Data Protection requirements. There are eight enforceable principles of good practice illustrated in

Figure 10.1.

Participants must be made aware that the research is happening, what will happen with the data and who it will be shared with. Additionally, if personal data has been obtained from them for a purpose other than research, they need to be asked their permission for their data to be used for research and be given the choice of opting out of the research. If their data is to be handed over to a third party for research to be undertaken, they, again, need to be asked for their permission and be given the choice of opting out and not being contacted.

Before undertaking any piece of research, it is recommended that you fully familiarise yourself with Data Protection requirements.

Figure 10.1 Data Protection Principles

The First Principle

- Personal data shall be processed fairly and lawfully.

The Second Principle

- Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.

The Third Principle

- Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed.

The Fourth Principle

- Personal data shall be accurate and, where necessary, kept up to date.

The Fifth Principle

- Personal data processed for any purpose or purposes shall not be kept longer than is necessary for that purpose or those purposes.

The Sixth Principle

- Personal data shall be processed in accordance with the rights of data subjects under this Act.

The Seventh Principle

- Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.

The Eighth Principle

- Personal data shall not be transferred to a country or territory outside the European Economic Area, unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.

10.2 ETHICAL ISSUES

The following **ethical considerations** need to be taken into serious consideration when undertaking any evaluation:

- Voluntary Participation;

- Informed Consent;
- No harm to participants;
- Confidentiality and anonymity;
- Analysis and reporting; and
- Exemption for research purposes

Ethical considerations will sometimes conflict with technical and practical considerations but it is important not to put technical and practical considerations above the interests of research participants.

10.2.1 VOLUNTARY PARTICIPATION

Participation in research should usually be voluntary. However, it is worth noting that voluntary participation can conflict with representative sampling since certain types of people are more likely to decline to take part in research than others. Having said this, compulsory participation is not the solution as it will usually undermine the quality of responses provided. It is best to encourage as many people as possible to participate voluntarily and to adjust for any known sampling biases that occur. This may involve undertaking additional research with groups that are under-represented.

10.2.2 INFORMED CONSENT

'Informed consent' is closely connected to voluntary participation. So that potential participants are able to make an informed choice about participation, the researcher should explain the information in Table 10.1 as clearly and as fully as possible.

Too much detail may confuse, overwhelm or distract potential participants rather than inform them. It is best to provide basic information and to offer to answer additional questions. Informed consent may be a process subject to renegotiation over time.

Informed consent can be gained verbally as well as by asking participants to sign a form. Signing consent forms can sometimes put people off taking part in research and make them more suspicious. In the event of children participating in Scottish Enterprise research, informed consent needs to be sought from their parents or guardians.

The respondent's right to privacy and the right to refuse certain questions, or to be interviewed at all, should always be protected and undue pressure should not be used to persuade people to take part in research.

Table 10.1 Informed Consent Checklist

Researcher Explanation of:	Check
• The purpose of the research and what it entails;	
• Who is undertaking and financing the research including the identity of the researcher;	
• An outline of any reasonably foreseeable risks, embarrassment or discomfort (for example some questions may illicit personal or sensitive information);	
• A description of the likely impact of the study;	
• A description of how the respondent was selected for the study;	
• A statement that participation is voluntary and that the respondent is free to withdraw at any time or to decline any particular question;	
• Details as to how the findings will be disseminated; and	
• An offer to answer any questions.	

10.2.3 NO HARM TO PARTICIPANTS

Research should not harm those who participate in it unless they willingly and knowingly accept the risks of harm. 'Harm' may mean revealing information about those being studied which might embarrass them or endanger their home life, friendships, jobs, status in the community and so on. The researcher must also guard against upsetting or psychologically harming participants, for example by them being asked to reveal personal attitudes or behaviour.

When surveys are administered through third parties, for example to employees via their employer, these third parties should not see the responses before they are returned to the researcher. To ensure that this does not happen, completed questionnaires should be placed in sealed envelopes and, ideally, collected by someone other than the employers.

10.2.4 CONFIDENTIALITY AND ANONYMITY

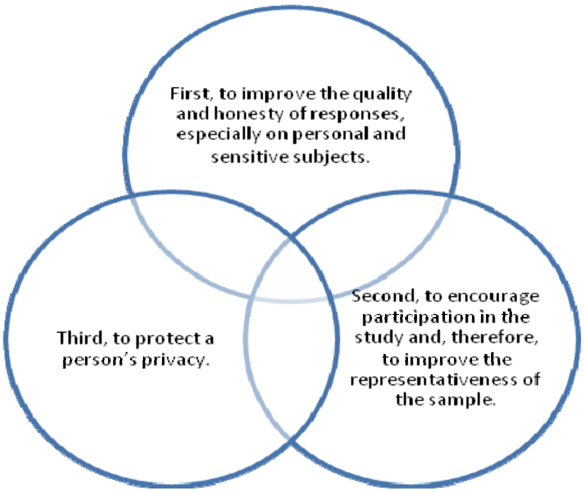
There are **three** main reasons for assuring confidentiality, as indicated in Figure 10.2. The most obvious way in which participants can be harmed in research is if the confidentiality of responses is not honoured. Usually, research participants are assured that their individual answers will either be anonymous or confidential. A respondent may be considered anonymous when the researcher does not, or can not, identify a given response with a given respondent. In a confidential survey, the researcher is able to identify a given person's responses but promises not to do so publicly. The information respondents provide should, in most cases, be kept confidential.

It is important not to promise confidentiality unless the promise can be kept. It is also important to be aware of the legal situation and to be aware that, in some cases, you may be required to pass on the information to a professional, e.g. fraudulent activity.

You can retain the confidentiality of data by keeping names away from responses and, for example, putting ID numbers rather than names on questionnaires. You may then need a separate file where the ID numbers are linked to names, particularly if follow-up of respondents is planned. Access to this file needs to be restricted.

As mentioned earlier, it is also important to ensure that confidential data does not get into the hands of third parties, for example questionnaires filled in by companies should not fall into the hands of their competitors.

Figure 10.2 Reasons for Confidentiality



10.2.5 ANALYSIS AND REPORTING

It is important not to falsify findings by being selective with the information that is presented or only presenting the positive or negative picture. To safeguard against this, it will be important to make data sets, such as SPSS files, available with the findings.

10.2.6 EXEMPTION FOR RESEARCH PURPOSES

Where personal data processed for research, statistical or historical purposes are not processed to support decisions affecting particular individuals or in such a way as likely to cause substantial damage or distress to any data subject such processing will not breach the Second Principle and the data may be retained indefinitely despite the Fifth Principle.

As long as the results of the research are not published in a form, which identifies any data subject, there is no right of subject access to the data.

11 MANAGING EVALUATION

11.1 THE COMMISSIONING PROCESS

The suitability of the proposed approach and methods to answer the questions asked in the Terms of Reference should be central to the selection of the preferred tender. The capabilities and experience of the team must be matched with the methodology proposed in order to avoid problems occurring whilst the evaluation is being implemented.

However there is a danger that this will discriminate against new entrants and therefore make the creation and/or maintenance of competitive markets more difficult.

A useful way to judge a proposed team is to ask to see previous examples of their work. This can be further supported by asking for ‘references’ from previous evaluation customers, i.e. named persons who can be consulted.

Finally, it is always good to pay attention not only to the presence in the team of highly qualified personnel, but also to the time that they are prepared to devote to the task. As evaluations are time consuming, the most qualified people will not undertake all the fieldwork themselves. Time allocated by those with experience needs to be sufficient to provide supervision for those working in the field. Evidence of the proposed team having worked together successfully is pertinent.

Figure 11.1 Contractor Management

Contractor Management

- Giving the consultants clear guidance as to the aims of the evaluation
- Keeping in regular contact (ideally weekly) through phone, email or face-to-face to discuss progress (a method should be agreed) and any issues/problems arising. It will be important to agree dates for project/programme reviews at the inception meeting
- Encouraging the consultants to discuss any problems they are facing as soon as they arise so that they can be dealt with speedily and do not hold up the evaluation unnecessarily
- Resolving issues and problems in partnership
- Building a good rapport with the consultants which allows you both to speak freely and openly encourages on going communication

11.2 APPROACHES TO MANAGING EVALUATION & INDEPENDENCE

Once the evaluation study has started there is the temptation for the commissioning authority to keep contact with the evaluation team at arms length. This view is based on the belief that a 'hands-off' approach will help to secure the independence of the evaluation team.

The independence of the team, in fact, depends on a much more complex set of factors than the mere reduction of contacts with the client. The best guarantee of independence is the scientific and professional standing of the selected team.

When managing evaluations commissioners and programme managers need to be aware that there are a number of ongoing factors that can undermine the independence of evaluators. All evaluation work requires a measure of independence between the evaluator and the object of evaluation. Even in self evaluation approaches those involved in the implementation of interventions need to achieve a degree of distance and independence whether or not they are assisted in the process by outside evaluators.

Normally increasing the level of independence of the evaluator from the object of evaluation will increase the credibility of the evaluation findings. In all circumstances the possibilities for conflicts of interest need to be minimised and where possible eliminated. Sometimes this is achieved through formal declarations from evaluators and potential evaluators as to the absence of such conflicts.

However, evaluators are rarely fully independent of the objects of evaluation and evaluation is never 'value free'. Evaluators will be subject to a whole range of 'influences'. Indeed the commitment of the evaluator to the aims of the intervention under consideration may well increase the quality of the evaluation findings and the chances that the results lead to improved economic development.

11.3 STEERING GROUPS

In many cases, it will be **helpful to set up a Steering Group** to guide the external consultants as well as providing a named lead contact. The principle role of the Steering Committee is to ensure a high quality and useful evaluation. This will involve facilitating the work of the evaluators through for example, providing access to information and contacts, and elaborating evaluating questions and key issues that they feel should be informed.

The Steering Committee should not attempt to influence the evaluators to omit certain evidence or to come to conclusions they would prefer to hear that are not substantiated by the evaluation evidence. The Steering Committee should also oversee the process of communicating the evaluation findings. Establishing an evaluation Steering Committee consisting of the different stakeholders in the programme makes it possible to guarantee:

- Better acceptance of the evaluation by those evaluated, by creating relations of trust;
- Easier access to information and a better understanding of the facts and events which took place while the programme was underway;

- Opportunities for 'process' use and learning among stakeholders as a result of their steering committee interactions;
- Interpretations and recommendations which take into account all the important points of view;
- The dissemination of conclusions and taking into account of recommendations more rapidly and informally; and
- A greater likelihood that recommendations and conclusions will lead to action and follow-up.

Generally, the Steering Committee should include four categories of people as indicated in Figure 11.2.

Figure 11.2 Steering Committee Membership

The strategic management of the programme or intervention,

- that is the funding authorities, the policy side of the administration and where appropriate the different levels of government.
- A multi-level approach to involving strategic management on the Steering Committee is very important as programmes grow increasingly complex taking into account concerns that have different territorial dimensions;

The operational management of theproject/ programme

- that is those whose activities are evaluated by the study, although in order to guard the impartiality of the Steering Committee, operational management is usually represented by senior managers, a little distant from the front-line of day-to-day management.
- Even so, it is an important task of Committee chairpersons to ensure that no members, including operational managers, attempt to influence evaluation findings or ignore any body of evidence;

The social partners

- that is, the people representing the main interests affected by the project / programme.
- These can include not only the beneficiaries, but may also extent to trade associations, trade unions and the economic interest associations.

The experts

- that is people that have either substantive or methodological knowledge that can be useful for defining the evaluation questions or interpreting the results.
- The presence of independent experts in the Steering Committee can be very important in providing useful inputs to the evaluation team and in order to open up debate towards the more general lessons that can and should be drawn from the exercise.

12 COMMUNICATING THE FINDINGS AND ENSURING ACTION

12.1 DISSEMINATING THE RESULTS

Communication is an important part of the evaluation process. Evaluation work is of no consequence unless the findings are communicated. It is better to treat the communication task as continuous: an opportunity for dialogue and the accumulation of understanding rather than put all communication efforts into one big dissemination exercise after a final report is delivered.

Communication should therefore include:

- Improving awareness of evaluation work that is underway;
- Providing feedback on interim findings;
- Circulating and managing feedback on draft reports and outputs (e.g. data collection instruments); and
- Communicating evaluation findings and conclusions (both in written format and through presentations or workshops).

It is important that evaluation findings and the recommendations are fed back into the project/programme delivery. This can be achieved through preparation of an action plan which specifies the actions to be undertaken, when and by whom. All of the recommendations from the evaluation may or may not be implemented but each should be considered and decisions on actions arising from them should be recorded.

The evaluation results should also be made available to a wider audience internally and, where appropriate, externally. This will allow them to share in the lessons learned from the evaluation. The report should be submitted for inclusion on Evaluations Online (http://www.evaluationsonline.org.uk/evaluations/eval/advanced.do) and made publicly available.

The report would be provided to other members of the Steering Group for the study. However, it may not be considered appropriate to disseminate the full report to a wider external audience. In this case a shorter report could be prepared for wider dissemination. Where this will be required it should be made explicit in the evaluation brief.

12.1.1 THE WRITTEN REPORT

The principal form of communication is a written report. Whilst the appropriateness of the particular means of communication will vary there are a number of good practice lessons, summarised below. When undertaking and writing up an evaluation it is essential to define the target audience for the evaluation at an early stage.

It is desirable to define the report structure early on in the process. It is also essential to prepare an executive summary (5-10 pages); it is more likely that this will be read in detail than the full report. Often preparing an early draft of the executive summary forces clarification of the key findings and recommendations.

Table 12.1 Reporting Good practice

Report Writing Good Practice
• The written report should be clearly written and concise. • One hundred pages including an executive summary are normally sufficient. • Detailed evaluative evidence such as case studies and quantitative analysis should be presented in annexes or made available separately.
• The report should include an executive summary of 5-10 pages written in a style suitable for policy makers.
• The links between the conclusions and the analysis of evidence should be clear.
• The drafting of the report should indicate the basis for the observations made: the evaluation evidence or a combination of evidence and the evaluators' opinion.
• The report should include a description and assessment of the method used that is sufficiently detailed and self critical to enable the reader to judge the weight of evidence informing the conclusions.
• Use should be made of tables and diagrams where they improve the presentation of findings.
• Reference should be made to good practice examples of interventions to illustrate the arguments being made but evaluation reports should not take the place of good practice guidance. • Pressure on evaluators to produce 'good news stories' is often counterproductive: such results are viewed with suspicion by public and policy makers alike.
• The recommendations made should be clear in the follow-up action that is required.

One common structure for the final evaluation report is:

- Executive summary;
- Introduction and background (e.g. identifying the strategic, market failure and equity rationales of the intervention; the detailed objectives; and including the state of knowledge in field, referring to other relevant SE evaluation and external evidence where available);
- Methodology and research tools (include copies of these in appendices (see Table 12.2));
- Findings (see Table 12.2);
- Discussion of findings with implications for future delivery and policy (including refinement of intervention objectives, contribution to GES objectives and targets)
- Conclusions;
- Recommendations;
- Sources; and
- Appendices (these will need to include the research tools and lists of contacts).

A main issue is therefore the documentation and presentation of findings. The presentation of findings should include the items listed in Table 12.2.

Table 12.2 Presentation of Research Results Checklist

Issue		Check
Purpose of Research		
	The purpose of the research; and	
	Organisations or individuals responsible for undertaking the work.	
Sampling Issues		
	The sampling frame (the population from which the sample was drawn), the sampling strategy and the size of response, particularly in relation to planned and actual responses;	
	Where appropriate, a statement of response rates, whether any responses have been weighted in the final results, and discussion of the any non-response bias in the results (i.e. whether the responses are representative of the population under study);	
Data Collection Methods		
	The time period over which the research was undertaken;	
	Details of how the research information was collected/obtained, including details of any quality assurance methods employed. Consistency with the Market Research Society Code of Conduct should be clearly stated; and	
	How respondents were contacted and persuaded to participate, including any incentives offered.	
Presentation of Findings		
	Findings should be presented in a manner suited to the needs of the intended audience(s).	
	A clear structure to the presentation of findings must be included, plain English used throughout and any technical terms explained with clarity for non-technical audiences;	
	Copies of survey instruments (such as questionnaires) should be included within the presentation of findings;	
	Clear distinctions must be made between factual responses, interpretation of results and recommendations arising from the research;	
	Where percentages are used to present findings, the base numbers from which these are calculated should also be clearly stated;	
	An indication of the probable accuracy of results should be presented, using formal margins of statistical error (confidence intervals) where appropriate; and	
	Where primary research results are ‘benchmarked’ against other studies, such as national surveys, a clear comparison must be made, incorporating vagaries of the benchmark results and their likely reliability. Such benchmarking should only be made when it is clearly relevant to the aims of the primary research and justifiable in terms of accurate comparisons.	

12.2 ENSURING ACTION AND CHANGE

What is the best way that an evaluation can lead to action? It is crucial that evaluation findings, and recommendations resulting from these, are fed back into the project/programme and policy development and planning processes. Undertaking evaluation work and ensuring its quality is only worthwhile if the activity leads to some use of the evaluation findings and contributes to improved knowledge amongst those best able to take advantage from it.

There are at least three different ways in which evaluation work is used:

- Even where evaluation results are not used the process of evaluation initiation and reflection can be useful by offering opportunities to exchange information, clarify thinking and develop frameworks.
- More often, several evaluations or individual evaluations combined with other evidence and opinion are used cumulatively to inform debates and influence decision-making. Evaluation work thus stimulates the process of debate, challenge and counter challenge to evidence and its interpretation.
- Individual evaluations may be used directly or in an ‘instrumental’ manner whereby the results, findings, conclusions and recommendations are taken up. In practice this is unusual and where it does occur it tends to take place only partially.

This best way to ensure that action results from an evaluation is for the final report to specify, very clearly, the recommended actions. These recommendations can be summarised in an **action plan** indicating what activities should be undertaken by whom and by when. This should be prepared in co-operation with project/programme managers. The action plan needs to be published and disseminated with the final evaluation report. The recommendations of the evaluation may, or may not, be implemented but it is important that they are carefully considered and that decisions arising from them are recorded.

Action plan should be signed off by the project or programme Senior Responsible Owner (SRO), and the plan included in the Quarterly Evidence Report. The action plan should be followed up six months later to assess progress towards its implementation.

This is the responsibility of project/programme managers as well as those closely involved in the policy and project/programme development and planning processes.

12.3 HINTS & TIPS

Table 12.3 Hints & Tips for Evaluation Design and Implementation

Evaluation Design and Implementation
1. Evaluation competence should be brought in early by programme planners. In particular, this can help clarifying objectives and the intervention logics of programmes. This activity, although employing evaluation competence, is quite separate from mainstream evaluation activities. It needs to occur at the programme design and planning stage. However, this can make subsequent evaluation easier and more successful. Various techniques such as evaluability assessment and preparing an analysis of 'programme theory' can be deployed for this purpose. In general, in order to ensure independence of the main evaluation, it would be best to use different evaluation teams or resources for this programme planning work than for the main evaluation.
2. A similar evaluability assessment should be undertaken by evaluators when they begin their work. To some extent this may overlap or repeat what has already taken place at programme planning stage. However, the purpose here is different. It is to ensure that a feasible evaluation plan is produced and to clarify how evaluation outputs will be used. This is consistent with a general expectation of evaluators that they should be concerned with how their results, conclusions and recommendations are used from the earliest possible stage of their work.
3. Stakeholders, programme managers and policy makers, potential beneficiaries and partners should be involved in the evaluation from the earliest stages, where practicable. This will ensure that the evaluation design and plan will include their priorities and agendas. It will also ensure that they feel some sense of ownership of the outputs of the evaluation and are more likely, therefore, to find these useful and use these outputs. On the other hand, it may be necessary to be selective in which 'voices' finally determine the evaluation agenda, in order to retain focus and ensure the evaluation is manageable. Overarching priorities should be shaped by the intentions and logic of the programme – whilst remaining open to unintended consequences especially for intended beneficiaries.
4. Evaluations need to be actively but sensitively managed. This will ensure that commissioners are aware of choices that need to be made along the way. It will also ensure that evaluators receive sufficient support, access to information and briefing as to changes in policy and context. Those responsible for commissioning an evaluation and programme managers are the most suitable people to manage the evaluation because they will be aware of its background and rationale.
5. It is usual to derive criteria for an evaluation, i.e. judgements as to the basis for positive and negative assessments of progress, from the objectives of a programme. It is also important to include a wider set of criteria that derive from social needs. For example, is this programme useful and helping those for whom it is intended? Does it support equity or not? Is the programme consistent with other policy initiatives? And, is it delivered in an efficient and legitimate way? Maintaining this broader perspective ensures that for part of their work at least, evaluators are able to stand outside the logic of the programme and take a critical perspective on what the programme is trying to achieve and how it does it.
6. The importance of evaluation questions in an evaluation design cannot be overstated. The temptation otherwise is to gather large quantities of data and produce sometimes technically sophisticated indicators which make little contribution to practice or policy. There is, of course, a problem formulating the evaluation questions in a way that they are likely to be able to be answered. Whilst this is a technical question, there is here also the overarching concern for use. You should try to ask questions that someone will find useful. However, use should not itself be defined too narrowly. We are talking here not just about the instrumental use of evaluation that managers have. We are also talking of uses that citizens and civil society groups may make of evaluation in support of democratic processes and accountability.
7. We have specified in some detail the content and form of an ideal Terms of Reference for an evaluation. This is part of the general question of design and the choices that can be made at the design stage which can influence the quality and direction of an entire evaluation. It is important therefore not to simply follow a standard framework with pre-drafted paragraphs. Rather it should be recognised that defining scope, clarifying the users of the evaluation and deciding the skills required for an evaluation team, are among the most important decisions that are made during the course of an evaluation.
8. It used to be common to regard the use of evaluation as being confined to acting on recommendations and final reports. It is now understood that evaluation use can be supported and occurs throughout an evaluation. So-called process or dialogue use should involve stakeholders in evaluation thinking from the beginning. There are even evaluations where the conclusions and recommendations are rejected but stakeholders, especially the core stakeholders involved in the steering committee, nonetheless find the evaluation useful. It can help them clarify their own thinking and understanding of the programme and spark off innovative ideas for programme improvement. This continuous process of communication provides a particular context for the dissemination of evaluation reports and findings. Promoting dialogue during the course of an evaluation is likely to ensure that when stakeholders receive reports they will be better prepared and receptive.
9. It is often easier for programme managers and those who commission an evaluation to confine judgements of evaluation quality to the outputs in reports of the evaluation itself. However, this quality control process provides few opportunities for learning and improvement in the way the evaluation itself is managed. A quality assurance perspective of the kind that has been advocated provides a context in which to explain the

Evaluation Design and Implementation
strengths and weaknesses of evaluation outputs. It also offers an opportunity for those who commission evaluations to learn how to improve evaluations in future.
10. Consideration should be given at an early stage to how evaluation findings will be put to use. Some use will stem directly from the findings and recommendations of the work. Individual evaluations can also be helpfully combined with other evidence to inform debates. The process of evaluation can bring benefits in terms of structuring inquiry and institutional reflection. Close attention to the factors that influence the use of evaluation work will maximise its contribution.

13 ACKNOWLEDGEMENTS & SOURCES

This guidance note adapts material from a number of sources (below). Our thanks also to colleagues from the Policy & Research Team, Professor Sean McDonald (Bentley University), and EKOS Economic Consultants for many useful comments and suggestions. Any errors or omissions are solely the responsibility of Scottish Enterprise.

13.1 SOURCES

Cabinet Office (2003) The Magenta Book, Guidance Notes for Policy Evaluation and Analysis, Cabinet Office: London

Cabinet Office (2004) Strategy Survival Guide, Cabinet Office: London

D.A. de Vaus (2002) Surveys in Social Research, 5th Edtn, Routledge: London.

European Commission (2003) The Evaluation of Socio-Economic Development, The Guide, December European Commission, Directorate General for Regional Policy: Brussels

European Commission (2006) SMART Innovation: A Practical Guide to Evaluating Innovation Programmes, Directorate General for Enterprise and Industry: Brussels.

Futureskills Scotland (2002) Conducting Primary Research, A Good Practice Guide in Primary Research into Labour Market and Economic Issues, Guidance Note 1, October, Futureskills Scotland: Glasgow

HM Treasury (2003) The Green Book, Appraisal and Evaluation in Central Government, Treasury Guidance, HM Treasury: London.

OECD (2008)OECD Framework for the Evaluation of SME and Entrepreneurship Policies and Programmes, OECD: Paris

Office of the Deputy Prime Minister (2004) Assessing the Impacts of Spatial Interventions: Regeneration, Renewal and Regional Development - 'The 3Rs guidance', ODPM: London.

Scottish Enterprise (1996) Project Development Guidance, June, Various Volumes, Scottish Enterprise: Glasgow.

Scottish Enterprise (2004) Appraisal & Evaluation Guidance, Scottish Enterprise: Glasgow.

Scottish Enterprise (Undated) A Positive Guide to Evaluation, Scottish Enterprise: Glasgow.

Welsh Assembly Government (undated) Step-by-Step Guide to Evaluation, Welsh Assembly Government: Cardiff

W. K. Kellogg Foundation (1998) Evaluation Handbook, W. K. Kellogg Foundation: Battle Creek, MI (www.wkkf.org).