

**BIS** | Department for Business  
Innovation & Skills

**GUIDANCE ON EVALUATING THE  
IMPACT OF INTERVENTIONS ON  
BUSINESS**

Annexes

AUGUST 2011

## Contents

Annex A: Logic model Small Loans for Business, Business Support Product .	1
Annex B: Five Step Process: using the logic model to ask questions to start scoping the feasible evaluation models .....	4
Annex C: BIS Economic Appraisal Checklist .....	6
Annex D: Cognitive testing of Annual Turnover question for a telephone survey. McGee et al. (2008) round 1.....	8
Annex E Analytical framework for a review of the drivers of business success and failure .....	12
Annex F Summary of evaluation theoretical models .....	13
Annex G: Checklist of issues to consider when setting up a survey .....	25
Annex H: Key threats to conducting a robust survey of businesses.....	36
Annex I: Some standard classifications and definitions .....	38
Annex J: ONS Virtual Microdata Laboratory .....	43

## Annex A: Logic model Small Loans for Business, Business Support Product

Narrative	Inputs	Activities	Outputs	Intermediate outcomes	Outcomes
Description of the product: Debt finance	RDAs identify funding for small loan provision within regions	Awareness raising with mainstream lenders and target audience	Investment in supported businesses increases.	Increase in growth of SMEs in target areas (deprived areas and minority ethnic groups).	Reduced income differentials between women and minority ethnic groups and industry as a whole.
To provide enterprise funding to entrepreneurs from disadvantaged groups or communities for viable business propositions unable to secure the finance from mainstream sources.		Provide loans to businesses in disadvantaged groups / communities.		Increase in start-up SMEs in target areas, amongst women and amongst defined minority ethnic groups.	

Narrative	Inputs	Activities	Outputs	Intermediate outcomes	Outcomes
		Advice to beneficiaries to succeed in business.			
	Total resources provided	Number of businesses (including target groups) receiving financial support	Total investment in target businesses. (not collected in BIS surveys or Business Rep surveys)	Survey of business sustainability of target groups improves.	Differential between target group GVA and industry GVA.
		Total value of loans provided to businesses in target groups.	Reduction in access to finance difficulties for disadvantaged groups (ASBS)	Reduction in defaults of businesses in target areas.	Target group turnover
			VAT registrations by	SME survival rates (BIS/ONS) by	Target group

Narrative	Inputs	Activities	Outputs	Intermediate outcomes	Outcomes
			deprived area	deprived area	employment
			Number of new women-owned businesses from Barclays Bank data (captures some but not all new businesses below the VAT threshold).	GVA of SMEs in target areas.	Change in GDP per head for target area and economy.
			Self-employment levels (LFS) by Ethnicity, gender, deprived areas.	Turnover SMEs in target areas	SME industry GVA
				Employment of SMEs in target areas	

## **Annex B: Five Step Process: using the logic model to ask questions to start scoping the feasible evaluation models**

---

### **Small Loans for Business logic model: Measurement of Intermediate outcome: increase in growth of SMEs in target areas (deprived and minority groups)**

---

#### **1. Intermediate outcomes and final outcomes defined. Are they measurable?**

To measure an increase in growth of SMEs in target areas (deprived and minority groups)

What indicators can we use to measure 'growth'? SMEs turnover, number of employees? What other indicators are there?

Is this information available in deprived areas and by minority ethnic groups and women? or do we need to collect baseline data, if so how? How much would this cost? What method/s would be appropriate?

#### **2. Target population:**

What do we mean by minority ethnic group? Measurable by director of company or number of ethnic group employees? Need to define what the working definitions would be, concentrate only on the larger ethnic groups?

What is the working definition of 'deprived areas' where the policy is going to be applied?

Small Loans for Business will be delivered by RDAs. Will all RDAs have the same working definition of 'SME' 'deprived' and 'minority groups' for this product?

Are there available datasets that would allow the evaluators to target these groups? If so how are these groups defined in these datasets?

#### **3. Mechanism:**

Define treatment. Is the treatment equal to all? If not what are the ramifications? Do we want to evaluate all of the implementation types or just the most common?

---

---

**4. Feasible impact evaluation models, look at the table of models available (Annex F):**

Do we have any available data on the outcomes? Can we put monitoring and evaluation systems in place?

Do we have before and after measures?

How to measure the counterfactual: Can the product be delivered at random? If not, is it possible to have a comparison group? What would be the strengths and weaknesses of this comparison group based on; matched sample? Random sample of businesses in England?

**5. Finally assess each of the options for evaluation worked out with strengths, weaknesses along side possible budget, timetable, ethical considerations and likely impact of the evaluation**

Random allocation and control are very expensive, ethically difficult and difficult in implementation as there may not be a complete list of deprived and minority groups SMEs; match sample and control also difficult and the success of the evaluation will be very dependent on the quality of the matching. Random comparison group cheaper but difficult if looking at minority groups business.

Does the cost of setting up the comparison group outweigh the benefits? How is the relationship policy cost/evaluation cost looking?

Other questions include perverse incentives in data monitoring – are delivery organisations the best placed to monitor progress?

---

## Annex C: BIS Economic Appraisal Checklist

---

### Green Book Requirements

---

- |  |   |
|--|---|
| <p><b>1 Does the business case provide a compelling case for government intervention?</b> [GB, p11-12]</p> | <ul style="list-style-type: none"><li>• What market failure rationale justifies government intervention?</li><li>• What evidence is provided to support the rationale?</li><li>• What have we learned from previous evaluations of this type of intervention?</li></ul>   |
| <p><b>2 Have SMART Objectives been set?</b> [GB, p13]</p>  | <ul style="list-style-type: none"><li>• Are the proposed objectives SMART? (<i>Specific, Measurable, Achievable, Relevant and Time-bound</i>)</li><li>• Are the proposed action/s consistent with the objectives?</li><li>• What changes in behaviour would be required to achieve the objective?</li><li>• Is there scope for unintended effects?</li></ul>  |
| <p><b>3 Does the business case set out a range of options?</b> [GB, p17]</p>                               | <ul style="list-style-type: none"><li>• Has the business case considered a range of realistic options to achieve the policy objective?</li><li>• Are there other possible approaches, which have not been considered, which may achieve similar effects?</li><li>• Does the business case include a do minimum option?</li></ul>  |
| <p><b>4 Have the various options been appraised?</b> [GB, p20-36]</p>                                      | <ul style="list-style-type: none"><li>• Have costs been valued for each option?</li><li>• Have benefits been identified for each option?</li><li>• Have costs been valued on an opportunity cost basis?</li><li>• Have sunk costs been excluded?</li><li>• Have benefits or costs been adjusted for reasonable estimates of additionality, displacement or substitution effects?</li><li>• Have future benefits and costs been discounted to present values using the appropriate discount rate?</li><li>• Have any necessary adjustments been made for the impact of taxes, subsidies and changes in relative prices between options?</li><li>• Have transfers been identified separately from efficiency gains or losses?</li><li>• Have unvalued impacts been considered (e.g. environmental impacts)?</li><li>• Does the business case consider, and value if possible, distributional impacts (if material or an objective of the intervention)?</li><li>• Have benefit and costs streams been adjusted for optimism bias?</li><li>• Have risks been treated correctly?</li><li>• Has the sensitivity of projected costs and benefits been tested against key assumptions?</li><li>• Has a preferred option been identified and justified?</li></ul> |
-

---

**5 Have suitable performance measures been identified?**

- Does the business case set out milestones and targets associated with expected outcomes and outputs?

**6 Are monitoring and evaluation arrangements in place? [GB p42-49]**

- Monitoring arrangements in place
  - Evidence and performance data will be collected
  - Evaluation Plan has been agreed
  - Review dates have been set
-

## Annex D: Cognitive testing of Annual Turnover question for a telephone survey. McGee et al. (2008) round 1

### Annual turnover Q4-5 (Q4-5b)

#### Question cognitively tested

"The next few questions are about the financial performance of your company. These questions are designed to enable BERR to assess the impact or likely impact of your participation in <scheme/grant>".

**Q4** "Firstly, can I ask, what is the current annual turnover of your business? You may also refer to turnover as income, sales, invoices or receipts"

£

(Don't know)

(Refused)

#### **If Don't Know:**

**Q5** "If you had to estimate your current annual turnover, roughly into which of the following bands would you put your business?"

#### **READ OUT**

£0	1
Less than £100,000	2
£100,000 - £500,000	3
£500,000 - £2million	4
£2million - £10million	5
£10million - £50million	6
More than £50million	7
(Don't know)	8
(Refused)	9

### Findings

Regardless of whether respondents had received a loan, grant or support, they were able to give answer Question 4 and provide what they felt to be an accurate one, often stating that it is part of their job to know what the company's annual turnover was.

The term 'current annual turnover' was understood as referring to:

- how much money has come into the business;
- the sales and gross turnover for the business;
- figure of sales;
- global turnover;
- VAT exclusive (i.e. not including the 17.5%); or,
- the revenue created in a year.

Not only was there variability between different companies, but some respondents felt that if another company worker had been answering the question, they would have thought of something different. For example one respondent thought of his answer in terms of sales, whereas his wife would have answered in terms of invoices.

Although the question asked about the 'current' annual turnover, respondents often thought of the time period of their last financial year, be it April to March, January to December, October to September, or June to May. Occasionally, respondents would think of the current financial year, although this tended to be in cases where the financial year was coming to an end. In the case of where the company was no longer trading, the respondent based his answer on the previous two years.

The annual turnover figure was easily accessible to respondents if an exact figure was needed, although generally respondents could provide off-the-top-of-their-heads, what they considered to be, a reasonably accurate estimate.

Among multi-site companies there was variation in the way in which answers were reported, with some providing an annual turnover figure for the site at which the respondent was based whilst others provided a figure for the company as a whole.

As respondents were able to answer Q4 with relative ease, no respondents in this test were asked Q5. However, one respondent mentioned when he read the question, that the figure bands were too broad.

#### **Q45 (Q4-5b) Recommendations for cognitive testing**

- Improve terminology, tailoring the question wording to each intervention.
- Add a question before Q4 which asks when the businesses financial year starts and finishes. This will help respondents think about this time frame when answering the questions.
- Change the wording of Q4 and Q5 to refer to a specific time frame (last complete financial year).
- Reduce the broadness of the bands in Q5 by adding in two extra categories (11 categories as opposed to nine). The decision surrounding the number of answer categories to use is often a compromise, trading off length of list versus level of accuracy. Wide answer bands can imply that precise data are not required and this was supported by the comments respondents gave during the cognitive testing.

---

**Proposed question(s)**

**Q4** *Please can you tell me when your business's financial year starts and finishes?*

*Interviewer: if financial year varies ask for the last complete financial year.*

INTERVIEWER: PLEASE CODE THE START MONTH.

January	01
February	02
March	03
April	04
May	05
June	06
July	07
August	08
September	09
October	10
November	11
December	12

INTERVIEWER: PLEASE CODE THE END MONTH.

January	01
February	02
March	03
April	04
May	05
June	06
July	07
August	08
September	09
October	10
November	11
December	12

The next few questions are about the financial performance of your business. These questions are designed to enable BERR to assess the impact or likely impact of your participation in <read out>.

**Q5a** What was the annual turnover of your business in the last complete financial year? You may also refer to turnover as income, sales, invoices or receipts.

£  
(DK or Ref)

If Q5a=DK

**Q5b** If you had to estimate the annual turnover of your business in the last complete financial year, roughly into which of the following bands would you place it?

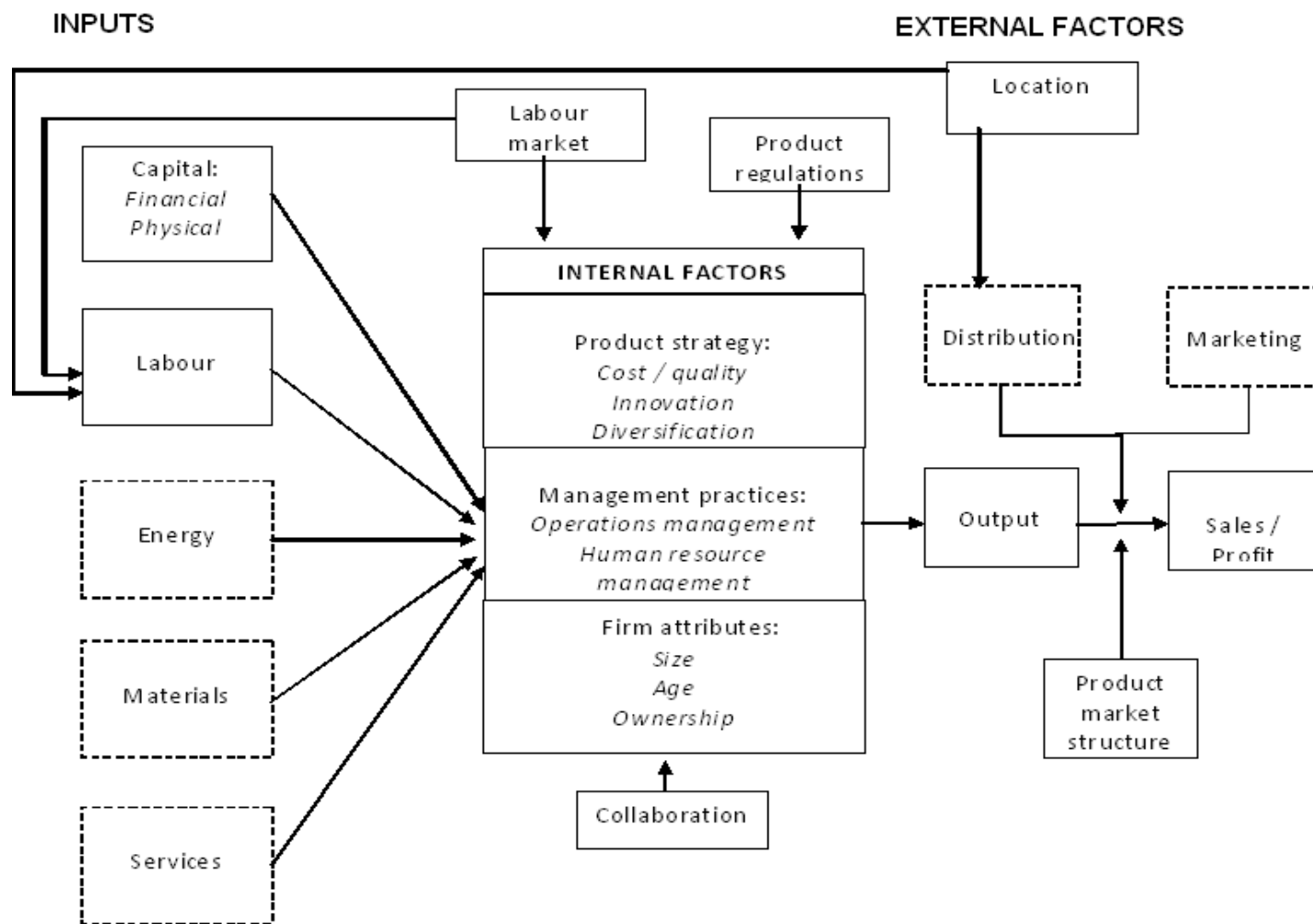
**READ OUT**

£0	01
Less than £100,000	02
£100,000 but less than £500,000	03
£500,000 but less than £1million	04
£1million but less than £2million	05
£2million but less than £5million	06
£5million but less than £10million	07
£10million but less than £25million	08
£25million but less than £50million	09
More than £50million	10
(Don't Know)	11
(Refused)	12

**Options for the Department to consider**

- Consider including a show card with the response options on them. This is discussed more fully in section 2.4.
- Consider the timing of the questions on economic impact in terms of how long after the intervention has come to an end should these questions be asked. This is discussed in detail in section 2.4.

## Annex E Analytical framework for a review of the drivers of business success and failure



Notes:

- Factors surrounded by a solid line are those covered in this review
- For simplicity, the schematic ignores feedback loops (e.g. from the internal organisation of the firm to investment decisions)

## Annex F Summary of evaluation theoretical models

The following table provides a summary of the most common evaluation models covered in Chapter 4. Although some might see this as an over simplified version of what is a very complex area it is important for evaluators to be familiar with the basic principles behind the models. **A fuller list of the challenges to validity and reliability and how these can enter the evaluation at each stage of the evaluation process can be found in Chapter 6.** An evaluation involves seeking to apply a theoretical model that seems most appropriate and adapting it to the particular circumstances. It is important to have the models in mind as they all have their assumptions, weaknesses and challenges to validity and reliability and capability to show impact or causality.

### Experimental or RCT:

#### Strong outcome based model to measure impact

Key characteristics	Model	Key strengths	Key challenges
<p><b>Randomised Control Trial (experimental/ control)</b></p> <p>There are always <b>two or more groups</b></p> <p><b>Random allocation</b> of treatment and control group</p> <p>There is a <b>before and after measure</b> of the treatment and control</p>	<p><b>Treatment group</b></p> <p>RO<sub>1</sub> X O<sub>2</sub></p> <p><b>Control group</b></p> <p>RO<sub>3</sub> O<sub>4</sub></p>	<p><b>Strong model to measure impact</b> for the purpose of testing hypothesized causal relationships.</p> <p><b>Random assignment</b> ensures that pre-experimental differences will be distributed more evenly among the groups or 'neutralised'.</p>	<p><b>Testing effects</b> can be a challenge particularly where businesses may be given a grant or some sort of assistance and they know they are being tested to evaluate it. The firm may also know whether it is in the treatment or control group. This can be reduced if the control group receive a lower level version of the treatment.</p>

Key characteristics	Model	Key strengths	Key challenges
<p>group</p> <p>It <b>assumes control</b> to avoid other confounding reasons for the effects</p>		<p>Maximum <b>control</b> of extraneous variables (alternative reasons that can have an effect on the outcome). The treatment and control group subjects experience the same events during the experiment except for the manipulation of the treatment.</p>	<p>It is difficult to fully control the environment of the treatment and controlled group. Alternative <b>reasons unknown</b> to the researcher could have an effect on the outcome undermining the validity of the findings. The effect seen could be due to a variable not measured in the study.</p>

## Quasi-experimental: Impact

There are always **two or more groups** but sometimes randomness is not possible or control is not possible. They are **outcome based models capable of providing robust estimates of impact**.

	Key characteristics	Model		Key strengths	Key challenges
Non-equivalent match or hypothetical comparison group design	<p><b>Matched group</b> treatment and comparison group by key characteristics</p> <p>There is a <b>before and after measure</b> of the treatment and control group</p> <p>It assumes <b>control</b> to avoid other confounding reasons for the effects</p>	<p><b>Treatment group</b></p> <p><b>Comparison group</b></p>	<p>O<sub>1</sub> X O<sub>2</sub></p> <p>O<sub>3</sub> O<sub>4</sub></p>	<p><b>Model capable of providing a robust estimate of impact, providing that the matching is done to high standard</b>, high percentage of matches by as many characteristics related to the outcomes as possible.</p> <p>The study will be more robust also depending on the <b>control</b> or measurement of extraneous variables (alternative reasons that can have an effect on the outcome). The treatment and the matched group</p>	<p><b>Relies heavily on the quality of the matched sample.</b> Difficult to obtain matched samples of good quality in the business community. Often a very small percentage of the target population is matched and the characteristics on which they are matched are not comprehensive enough. Or the matching technique only matches the most 'common' cases and fails to match those that are more on the extremes of the normal curve.</p>

Key characteristics	Model	Key strengths	Key challenges
		<p>subjects should experience the same events during the experiment except for the manipulation of the treatment.</p>	<p>It is difficult to fully control the environment of the treatment and matched group. Alternative <b>reasons unknown</b> to the researcher could have an effect on the outcome undermining the validity of the findings. The effect seen could be due to a variable not measured in the study.</p> <p>Testing effects as per experimental.</p>

Key characteristics	Model	Key strengths	Key challenges	
Multiple time series design: longitudinal comparisons	<p>The initiative has not been applied to all the population therefore it is possible to have a <b>comparison group</b></p>	<p>Treatment group    O<sub>1</sub> O<sub>2</sub> O<sub>3</sub> O<sub>4</sub>    X    O<sub>5</sub> O<sub>6</sub> O<sub>7</sub> O<sub>8</sub></p>	<p>See non-equivalent match or comparison group design. <b>Model capable of providing a robust estimate of impact.</b></p>	<p>See non-equivalent match or comparison group design.</p>
<p>If the comparison group is matched it would provide more robust findings</p>	<p>Comparison group    O<sub>9</sub> O<sub>10</sub> O<sub>11</sub> O<sub>12</sub>    O<sub>43</sub> O<sub>14</sub> O<sub>15</sub> O<sub>16</sub></p>	<p><b>A matched comparison group</b> will make it a stronger design to prove causality providing that the matching is done to high standard.</p>	<p>If a matched group is not possible the estimate of impact will be dependent on the quality of <b>the comparison group.</b></p>	
<p>There are many <b>before and after</b> observations on the treatment and comparison group</p>				
<p>It assumes</p>				

Key characteristics	Model	Key strengths	Key challenges
<b>control</b> to avoid other confounding reasons for the effects		Having <b>several measures or observations</b> before and after should make the model stronger.	

### Quasi-experimental: other outcome based models

There may be two or more groups, sometimes just samples of the same target population, sometimes randomness or control is not possible, before treatment observations may not be available. They are not really capable of measuring impact (prove causality of the policy or intervention) but are **useful outcome based evaluation models** which allow evaluators to see if the intended outcomes or targets of the intervention are being met, whether they are the direct outcome of the intervention or not. **These models cannot robustly measure impact/causality but can look at strong correlations and associations.**

Key characteristics	Model	Key strengths	Key challenges
<p>Separate sample Pre-test/ post-test design</p> <p>Different random samples from the same population</p> <p>Before measure of one random group and after measure of a different random group of the same population that is know to have received the treatment</p>	<p>Treatment group (sample 1) <math>RO_1</math> X</p> <p>Treatment group (sample 2) X <math>RO_2</math></p>	<p><b>Model capable of providing estimates of outcomes.</b></p> <p><b>Before and after measure in different samples</b></p> <p><b>Random assignment</b> ensures that pre treatment differences will be distributed more evenly among the groups or 'neutralised'</p> <p>Good to use when the <b>whole population</b> is being treated.</p>	<p>Can not prove causality of the intervention but allows evaluators to see if the intended outcomes or targets of the intervention are being met. Does not measure what would have happened anyway.</p>

Key characteristics	Model	Key strengths	Key challenges	
<p>Post-test only control group design</p>	<p>The initiative has not been applied to all the population</p> <p>The treatment and comparison group are chosen at random</p> <p>There is no before measure</p>	<p><b>Treatment group</b>      R   X   O<sub>1</sub></p> <p><b>Comparison group</b>      R      O<sub>2</sub></p>	<p><b>Model capable of providing estimates of outcomes.</b></p>	<p>Can not prove causality of the intervention but allows evaluators to see if the intended outcomes or targets of the intervention are being met. No observation before the intervention means there is no baseline for the treatment or control groups. Thus not being able to calculate whether there were differences in the groups before the intervention.</p>

Key characteristics	Model	Key strengths	Key challenges	
Static group comparison	Treatment is not at random	<b>Treatment group</b> X O <sub>1</sub>	<b>Model capable of providing estimates of outcomes.</b>	<b>Relies heavily on the quality of the matched sample.</b>
	A comparison group is chosen, ideally a matched group	<b>Comparison group</b> O <sub>2</sub>	<b>Stronger model with good matched groups</b>	Can not prove causality of the intervention but allows evaluators to see if the intended outcomes or targets of the intervention are being met. No observation before the intervention means there is no baseline for the treatment or control groups. Thus not being able to calculate whether there were differences in the groups before the intervention.

## Pre experimental designs

Observations possible only on one group. **Useful outcome based evaluation models.**

	<b>Key characteristics</b>	<b>Model</b>		<b>Key strengths</b>	<b>Key challenges</b>
Time series	<p>Can also be done following a random sample throughout time or with data of the whole population</p> <p>Many before and many after observations</p>	<p><b>Treatment group</b></p>	<p>O<sub>1</sub> O<sub>2</sub> O<sub>3</sub> O<sub>4</sub>    X    O<sub>5</sub> O<sub>6</sub> O<sub>7</sub> O<sub>8</sub></p>	<p><b>Model capable of providing estimates of outcomes.</b></p> <p>Where there is no possibility of having a control group or the treatment has been given to the whole population. The trend will give an indication of whether there is a significant change after the treatment from before.</p>	<p>Can not prove causality of the intervention but allows evaluators to see if the intended outcomes or targets of the intervention are being met.</p> <p>If a significant change is observed after the intervention it cannot explain whether the change would have happened anyway.</p>
One group pre-test/post-test design		<p><b>Treatment group</b></p>	<p>O<sub>1</sub>    X    O<sub>2</sub></p>	<p><b>Model capable of providing estimates of outcomes.</b></p>	<p>Can not prove causality of the intervention but allows evaluators to see if the intended outcomes or targets of the intervention are being met.</p>

---

One shot case study

Self assessment surveys often used

**Treatment group**

X O<sub>1</sub>

**Model capable of providing estimates of outcomes.**

Can not prove causality of the intervention but allows evaluators to see if the intended outcomes or targets of the intervention are being met.

Self assessment surveys are an example of 'one shot case study' where beneficiaries or recipients of an intervention are asked to provide measures of impact of the intervention. These provide a measure of perceived impact by the beneficiary or recipient of intervention, it can not be said to be actual impact and it is advisable to triangulate this with other methods.

### **Note on statistical techniques:**

See further reading sections for publications on statistical techniques and their application in different settings. The following summary outlines some commonly used techniques:

**Sampling:** Probability sampling (random assignment)

For **looking at simple before and after differences and associations** standard statistical techniques can be used in most models above, to name a few: simple and multiple regression analysis; t-tests

For **modelling programme or intervention effect** often use: difference-in-differences; regression discontinuity design; instrumental variables, longitudinal or time-series analysis.

For **matching treatment and control group** can use: propensity score matching, nearest neighbour, matching by profile characteristics.

---

## Annex G: Checklist of issues to consider when setting up a survey

A set of general criteria were developed by NatCen to evaluate three BIS surveys. These criteria can also be seen as reflecting best practice in developing and implementing surveys and could be used to evaluate any survey. They could also act as a useful checklist for those involved in the early stages of commissioning and contracting. In addition to these general criteria a set of specific criteria were developed, to assess whether each survey fulfilled its objective of measuring the economic impact of the intervention. The following tables display both sets of criteria, along with an accompanying full description.

### *General criteria*

Criteria	Description
<b>1. Research aims</b>	
a) Measurement objectives clearly set out	The aims underpinning the research need to be clearly formulated and communicated. The objectives should be easy to understand, short, succinct and to the point. It needs to be clear exactly what the research is intended to <b>measure</b> . A supporting rationale for why the data are being sought and how they are intended to be used assists in communicating these clearly. Furthermore, the objectives need to be achievable in the sense that they are realistic both in terms of timing and the nature of the work being undertaken.
b) Level of accuracy of data specified	The level of accuracy required for survey data must be clarified at the beginning of the research process, as this will influence how the data are to be collected, sample size and question construction. It may be that a general observation of a trend or incidence may be sufficient or that a much more precise figure may be required.

Criteria	Description
c) Research team well informed about the topic area	The person leading the research should be well informed on the topic area and may wish to put together a <b>steering group</b> to assist in shaping the research and ensuring it remains on the right track. A <b>feasibility test</b> or small scale <b>pilot</b> test may be useful in helping ensure the survey design is robust.
<b>2. Definitions</b>	
a) Key concepts defined	Key survey concepts should be fully defined at an early stage in the project (e.g. spelling out precisely what is meant, for example, by 'effectiveness' or 'impact'). Such definitions are important in operationalising the key concepts and ensuring consistency and continuity of measurement.
b) Concepts communicated effectively to target population	The concepts will need to be communicated to respondents in language they are familiar with. It is important to allow sufficient time for question design and testing in the survey timetable.
<b>3. Target population</b>	
a) Sampling unit(s) defined	The survey population needs to be clearly defined, specifically who or what will constitute the sampling unit and thus the unit of analysis (e.g. if the sampling unit is businesses this needs to be defined, for example, in terms of sites, regional offices, head office etc or the organisation as a whole).

Criteria	Description
b) Sampling methods appropriate to meet aims	Surveys requiring statistical reliability will require a random probability sampling design. The design should seek to minimise sampling error.
c) Sample size and composition sufficient to meet aims	The overall size of the sample is important and should be informed by considerations about the kinds of analyses that will be conducted and the population groups of interest (e.g. if doing analyses by type of business it may be necessary to boost certain groups expected to yield small base sizes). It may also be necessary to stratify the sample by the groups of interest (e.g. area, size of business).
d) Sampling frame(s) used ensure(s) good population coverage	The sample frame must be as complete and robust as possible and contain all information necessary to stratify the sample to ensure sampling and coverage errors are kept to a minimum.
<b>4. Data collecting mode</b>	
a) Mode maximises survey participation	The data collection mode(s) selected should ensure the possibility that as many businesses as possible are willing and able to participate. Using a mode that excludes or restricts particular groups' participation in the survey will introduce response bias, effecting the representativeness of the survey.

Criteria	Description
b) Maximises data quality	The pros and cons of different data collection modes (interviewer administered, self-completion, computer assisted, paper etc) should be assessed to highlight any potential problems with the quality of information being obtained. Factors that can influence quality of data obtained using different modes include: the sensitivity of the subject matter; the amount of information to be collected; the complexity of the information being sought; and the nature of information being sought (e.g. the amount of open questions, whether the question order is important). There may be some tension between this criterion and 4a) above.
c) Maximises chances of respondents being able to provide information being sought	It may be appropriate to send respondents information in advance of the interview, depending on the nature of the topic and survey mode. This is particularly relevant if the information needs to be looked up (e.g. financial records). The benefits and drawbacks of doing so need to be taken into account before a final decision is made.
<b>5. Questionnaire design</b>	
a) Questionnaire content established	The <b>content and priorities</b> of the questionnaire should be decided after re-viewing the survey aims. A period of <b>developmental work</b> may precede the question design phase, to scope out the important issues. This may be through consultation with experts in the topic area and/or recipients of a certain product or policy. The questionnaire must be kept within an agreed time limit to ensure respondent burden is kept to a minimum.

Criteria	Description
b) Most appropriate respondent(s) identified	<p>Depending on the nature of the survey it may be necessary to establish the person most suited to participate in it. For example, in a survey of businesses there may be one person to target or the information required may be spread across a number of individuals. Development work can be useful in identifying whether respondents have the information that is being sought and in defining who should be targeted as the 'main' respondent.</p>
c) Questions designed that respondents understand, are willing and able to answer	<p>The questionnaire must be designed to enable all types of respondent to easily provide the information required and ensure results are comparable (e.g. the characteristics of recipients of a particular product or scheme may differ widely and therefore the questions must make sense across the board). Some basic principles of questionnaire design include: break down complex tasks into 'bite size' pieces; use clear, simple syntax; use clearly defined terms and concepts (see 2a); request information that respondents will be able to provide; use easily accessible time scales and reference periods; route questions so respondents answer questions only relevant to their situation; and ask questions that are deemed appropriate and acceptable. Investing time in questionnaire design is vital to the success of the survey.</p>

Criteria	Description
d) Questionnaire fully tested	The questionnaire must be subjected to <b>rigorous testing</b> to ensure it is fit for purpose and that measurement error is kept to a minimum. Respondents can experience problems at each of the following stages of answering a question: comprehension, retrieval of information, judgement and response <sup>1</sup> . When filling in a self-completion form respondents may also experience problems in navigating through the document, particularly with routing to relevant questions.
e) Pilot test conducted	With new surveys it may be necessary to conduct a pilot test to ensure all aspects of the survey design are working as effectively as planned. A pilot test will uncover problems with the process, such as unwillingness to take part, difficulty in pinpointing the correct person to take part or high item non-response at certain questions. It will also provide evidence on the length of interview (if the questionnaire is to be administered by an interviewer). The pilot may also identify problems with the proposed systems for the returning of questionnaires or editing of data, for example.

---

<sup>1</sup> Collins, D (2003) 'Pretesting survey instruments: An overview of cognitive methods' Quality of Life Research 12:229-238.

Criteria	Description
<b>6. Data collection</b>	
a) Clear guidance and training for interviewers/fieldworkers provided to ensure that data are collected in a consistent way and interviewer bias minimised	Interviewers will need to be very clear about the survey's objectives and their role in the data collection process. They should be fully briefed shortly before fieldwork and provided with written documentation to refer to during it. Interviewers should have the necessary skills to encourage potential respondents to take part and reassure them that the information they give is important. There should be rigorous quality control measures in place to ensure interviewer error is minimised.
<b>7. Data processing</b>	
a) Coding and editing procedures documented, staff trained and quality monitored	Data cleaning and editing can be conducted in different ways, so once decisions have been made on how this will occur these need to be documented. This process must be carried out by trained staff, in a systematic way, to ensure processing error does not creep in. Quality control procedures should be in place, and the unedited data should be retained (for an agreed period).
b) Inconsistencies in the data are checked with respondents	When data appear to be inconsistent or findings are unexpected, these data will need to be investigated to uncover exactly why this may be. Ideally such inconsistencies and unexpected answers (i.e. values outside the expected normal range) should be reconciled with the respondent.

Criteria	Description
c) Missing values and outliers treated consistently	It must be clear how missing values and outliers will be treated and this should be documented.
d) Data manipulation procedures (e.g. derived variable creation) documented allowing replication	Derived variables (variables created by the analyst that involve combining information from two or more sources) should be documented to allow others to replicate their creation, and to facilitate in checking that they have been derived correctly.
<b>8. Data analysis</b>	
a) Errors/biases dealt with consistently	Survey bias and errors must be dealt with consistently to ensure comparability across responses is retained. Decisions need to be made about how to deal with outliers and non-response when analysing the data. Measurement error and the effect it has on the data must also be considered in interpreting findings.
b) Data analysed using appropriate statistical methods and model assumptions fully documented	Where appropriate, the resulting data should be comparable with statistics from other surveys within the socio-economic domain. This point should also be considered when thinking about mode of data collection, as mode can affect comparability of survey findings <sup>2</sup> .

---

<sup>2</sup> de Leeuw E.D. (2005) To Mix or Not to Mix Data Collection Modes in Surveys. Journal of Official Statistics, 21 (2): 233-255.

Criteria	Description
<b>9. Outputs and dissemination</b>	
a) Reports are clear, well-written, timely and accessible	The report should clearly address the research objectives. All outputs should be accessible in terms of being easy to locate, read, interpret and understand. The report should be produced within a reasonable time frame following the fieldwork period and delivered on a date agreed well in advance.
<b>10. Management issues</b>	
a) Successful commissioning process	The commissioning process should aim to follow the Social Research Association guidelines (pages 6 to 8) <sup>3</sup> wherever possible. Those who are managing the commissioning process internally must be able to reflect both policy and research expertise. The Invitation to Tender (ITT) is required to set out clear aims for the research and requirements for outputs. The tendering process should also be open to a wide range of contractors.

---

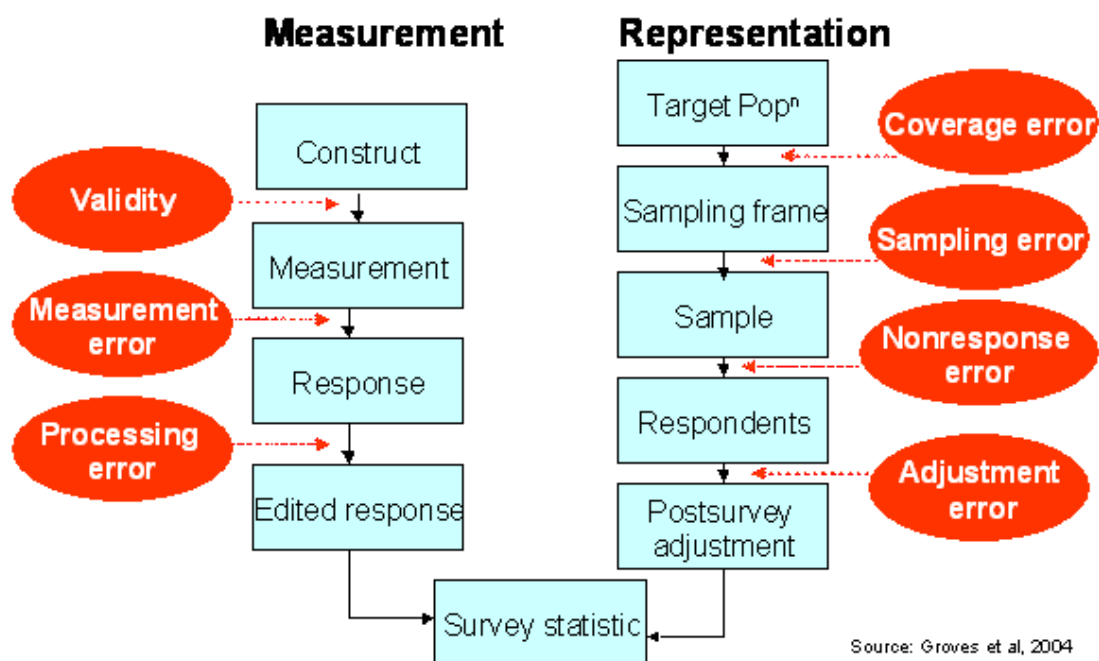
<sup>3</sup> The SRA guidelines can be found at [http://www.the-sra.org.uk/commissioning\\_sr.htm](http://www.the-sra.org.uk/commissioning_sr.htm)

Criteria	Description
b) Effective dialogue with research contractor established	It must be clear who the project manager will be and what their responsibilities will entail – both on the client and contractor sides. As stated earlier, it is important that the contractor understands exactly what data are required to address the key research objectives. How policymakers and customers will use the findings should also be communicated. The attendance of research commissioners/ project managers at interviewer debriefings and/or listening in on interviews is extremely useful in ensuring quality control and assessing how successful the survey design is.
c) Quality control monitored	There must be verification that the research instrument is 'fit for purpose' based on findings from piloting and early analyses of data collected.
d) Timetable adhered to	The timetable should be monitored to ensure that the data are produced on time, and in a coherent and appropriate format, so they can be of maximum value to the policy customers.
e) 'Value for money' of survey assessed	Once the project is completed a decision should be made regarding whether it provided 'value for money' and thus if it were to be repeated how improvements could be made. This knowledge should be documented so that it can then feed into future ITTs to help produce high quality and cost-effective work.

*Specific criteria evaluating the measurement of economic impact through self assessment*

Criteria	Description
11. Evaluating economic impact	
a) Reliable self-reports of economic change achieved	Respondents must be willing and able to give robust, accurate details of the way in which particular policies, initiatives or schemes affect their company. The questions must be designed so that they are easy to understand and can be answered using information that one might reasonably expect the respondent to have access to.
b) Respondents able to construct the counterfactual (i.e. change without the policy/initiative as an estimate of total change)	In addition to giving full information regarding the <b>actual</b> change the policy or initiative had on the company, respondents must be fully equipped to be able to consider the <b>counterfactual</b> : what would have happened had they not participated in or been affected by the programme, scheme or policy? Respondents can be asked to construct the counterfactual when they have not received the treatment (for example, their grant application was not successful). The counterfactual in this case is what they think would have happened had they received the treatment. In both cases respondents are asked to quantify the impact: to say by how much things would have been different. There should be internal consistency in the story being told, i.e. data on economic change should be consistent with the respondent's overall assessment of the counterfactual.
c) Low proportions of item missing at economic impact questions	The questions on economic impact should yield full, high quality data with few missing responses.

## Annex H: Key threats to conducting a robust survey of businesses



### Representation

Coverage error	The list of the population of interest is not full and has systematic errors, for example, it might miss all of the small businesses or businesses in a certain geographical area
Sampling error	Non-probability (non-random) samples can cause bias
Non-response error	Non-respondents can have systematic differences from the respondents
Adjustment error	Where weighting is applied but it is not done properly

### Measurement error

Construct error	Questions that are not clear or badly phrased will introduce error. Question development is very important and specialists survey methodologists should ideally develop the questions so not to fall into elephant traps. See Annex G for an example of the cognitive process on a question that just asks businesses about their turnover. See McGee et al. (2009) for more detail on cognitive testing.
Measurement error	Using the wrong interview mode, for example a business may not want to give financial data face-to-face, particularly if they have not been doing well.

Not screening properly for the right respondents, for example, when cold calling a business, the interviewer may end up asking questions about business performance to the receptionist.

Interviewer systematically making the same mistake when asking a question.

Response

Respondents answer questions in the way they think the interviewer wants them to answer rather than according to their true belief for example. This can further split out into:

Strategic bias      Response anticipates potential use of information and tries to maximise self-benefit.

Cognitive bias      It is costly to gather and process the information required and proxy responses are provided instead.

Interviewer bias      Respondent provides answer she thinks will gain approval from interviewer.

Warm glow bias      When asked how much they would be willing to pay for a public good, respondents don't assess the magnitude of the impact but just the opportunity to contribute to a good cause.

Scope bias      Respondents fail to assess the magnitude of services they are asked to value.

Property right bias      If respondents feel entitled to it, they would refuse to provide measures of willingness to pay for a good or service.

Processing errors

Coders can make a systematic mistake when coding the answers to a question. Analysts can run the wrong filtering or analysis.

## **Annex I: Some standard classifications and definitions**

### **Enterprise**

An Enterprise is the smallest combination of legal units (generally based on VAT and/or PAYE records) which has a certain degree of autonomy within an Enterprise Group. It may comprise a number of local units.

### **Local unit / Establishment**

Local units (sometimes referred to as establishments, especially in the US) are the local presence of an enterprise. So, for example, while a supermarket chain may form a single enterprise, it may have many local units. The head office would also count as a local unit.

### **Registered / unregistered**

It may be important to consider whether businesses are registered or not. Typically, this will refer to being registered for VAT and/or PAYE and this is what is held on the ONS Inter Departmental Business Register (IDBR). However, depending on the data source being used, it could refer to just one of these, or it could refer to being registered with Companies House (see below under legal status)

### **Employees / employment**

The number of employees refers to the number of people working within the enterprise under a contract of employment in return for a wage or salary. An enterprise can be classed as having no employees if all the business is conducted by people classed as being working proprietors.

Employment refers to the number of employees plus the number of self-employed people (working proprietors) who run the enterprise. Individuals with two jobs in different enterprises, and self-employed people running two enterprises will be counted against each enterprise. Both full-time and part-time employees are counted, and both are counted as an employee.

### **Data linkage & consent**

It is useful to be able to link data on businesses to administrative records, for example those held on the IDBR. To do this requires:

- i) A variable that can be used to link – most likely to be the VAT registration number, but could also be Companies House registration number;
- ii) Consent to carry out the linkage – to ensure compliance with data protection laws. An example of a form of words that could be used (taken from the BIS Small Business Survey) is:

*Would it be possible for BIS to link your responses to other information that you have provided previously to the Government? By this data linkage, we can reduce the burden of our surveys on your business and can improve the evidence that we use. Data will only be used to inform research on businesses in aggregate - we will never release information that identifies any individual business - and your survey responses remain strictly confidential. Do you give your consent for us to do this?*

It may also be a good idea to enable future contact with a business for a follow up evaluation. Again, a suitable form of words is necessary to make this possible. An example is: The Department for Business Innovation and Skills (BIS) may want to carry out further research in the future. Would you be willing to help with that research?

### **Majority led business**

Where details of owners are collected (see below), this can enable analyses of businesses that are majority owned by certain groups. There are two standard groups in which there is policy interest although note that the definitions are slightly different:

- Majority women-led business – a business where more than 50% of the people in the management team are women (designed to exclude businesses with husband / wife etc.)
- Majority MEG-led business – a business where 50% or more of the people in the management team are from a MEG (Minority Ethnic Group)

### **Standard categories**

Size bands (names)

Often there is interest in a particular type of business. Although there are different definitions for how to define business size, the one that BIS uses most often is based on the number of employees:

- Micro – with 0-9 employees
- Small – with 10-49 employees
- Medium – with 50-249 employees
- Large – with 250+ employees

## **Size bands (number of employees)**

It may be appropriate to use a more detailed version of the above. The version used by BIS in estimating the number of businesses (and therefore useful for being able to gross up) is below although it may be appropriate to group some sizes together, for example if sample sizes do not allow such detailed disaggregation:

0, 1, 2-4, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500+

## **Turnover bands**

Because business size tends to be defined by the number of employees rather than turnover, there is less standardization here. One classification used in analysis of our Small Business Surveys is:

- Up to the VAT threshold (£73k in 2011-12)
- From VAT threshold to £100k
- £100k - £249k
- £250k - £499k
- £500k - £999k
- £1m - £1.49m
- £1.5m – 2.8m

An alternative, used by ONS is:

- £0 - £49k
- £50k - £99k
- £100k - £249k
- £250k – £499k
- £500k - £999k
- £1m - £4.99m
- £5m+

Any survey dealing predominantly with larger businesses would need to disaggregate the larger categories more, but there is no standard banding used.

## **Industrial sector**

The standard classification that should be used where possible is SIC 2007. This is a hierarchical structure comprising 21 high level sections; 88 more detailed divisions and more detailed levels below that. More details at: [http://www.statistics.gov.uk/methods\\_quality/sic/downloads/SIC2007explanatorynotes.pdf](http://www.statistics.gov.uk/methods_quality/sic/downloads/SIC2007explanatorynotes.pdf)

## **Legal status**

There are 3 main legal forms that a business can take:

- Companies - or incorporated businesses are businesses that are registered at Companies House. All companies should be on the IDBR.
- Sole Proprietor – A business with a single owner which is not incorporated – this covers the majority of very small businesses, most of which will comprise a single self employed owner
- Partnership – As sole proprietors, but with more than one owner

## **Geographical disaggregation**

If data can be disaggregated, the most likely categories to use are region / country within the UK:

- East of England
- East Midlands
- London
- South East
- South West
- North East
- North West
- West Midlands
- Yorkshire and the Humber
- Northern Ireland
- Scotland
- Wales

In future, there may be less emphasis on regions, and other areas may be more appropriate.

## **Age of business**

It is sometimes useful to know how long a business has been operating (which would include any changes in ownership or legal form). Typically, BIS would use:

- Less than 1 year
- Between 1-2 years
- Between 2-3 years
- Between 3-4 years
- Between 4-5 years
- Between 6-10 years
- Between 11-20 years
- Over 20 years

## **Some definitions from Business link for business less than 1 year old:**

### **i) Pre-start**

A customer not trading yet but enquiring about or trying to start up. The count will be based on the 'status' of the start up business at the point the 'count' is reported.

### **ii) Start-up**

A business customer that has begun trading and is less than 12 months old. The start date is the date, deemed by the client, of first transaction. If the business age is not known, please add the business to "established businesses", and provide commentary if there are more than 10% unknown. The count will be based on the 'status' of the start up business at the point the 'count' is reported.

### **iii) Established business**

A business customer that began trading more than 12 months ago. The count will be based on the 'status' of the established business at the point the 'count' is reported.

## **Details of ownership**

If details about a business owner (or owners) are recorded, there are several dimensions that it would be worth considering. Eg:

- Gender – M/F
- Ethnicity – MEG / not
- Disability – DDA disabled / not
- Age – one possibility: <25, 25-34, 35-44, 45-49, 50-54, 55-64, 65+
- Qualifications – postgraduate (Level 5); degree / equivalent (Level 4); A levels / equivalent (Level 3); O levels, GCSE / equivalent (Level 2); Other; None

## Annex J: ONS Virtual Microdata Laboratory

### Internet address:

<http://www.ons.gov.uk/about/who-we-are/our-services/vml/about-the-vml/vml-overview>

### Example of some of the datasets it includes

Annual Inquiry into Foreign Direct Investment - AFDI  
Annual Population Survey - APS  
Annual Survey of Hours and Earnings - ASHE  
Annual Responders Database - ARD  
Business Enterprise Research and Development - BERD  
Business Spending on Capital Investment - BSCI  
Business Structure Database - BSD  
Capital Stock  
Census  
Community Innovation Survey - CIS  
DECC R & D  
E-Commerce  
General Household Survey - GHS  
Investments in Intangible Assets - IIA  
International Trade in Services ITIS  
Labour Force Survey - LFS  
Monthly Inquiry into Distributive and Service Sectors - MIDSS  
Monthly Wages and Salaries Survey - MWSS  
Monthly Production Inquiry - MPI  
New Earnings Survey - NES  
Employers Skills Survey - NESS/ESS  
New Earnings Survey Panel - NESPD  
Occupational Pension Scheme Survey - OPSS  
Prices  
Prodcom – Products of the European Community  
Quarterly Fuels Inquiry - QFI/QFD  
Quarterly Capital Expenditure Survey - QCES  
Scottish Employment Data  
Scottish Employer Skills Survey

© Crown copyright 2011

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. Visit [www.nationalarchives.gov.uk/doc/open-government-licence](http://www.nationalarchives.gov.uk/doc/open-government-licence), write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk).

This publication is also available on our website at [www.bis.gov.uk](http://www.bis.gov.uk)

Any enquiries regarding this publication should be sent to:  
Department for Business, Innovation and Skills  
1 Victoria Street  
London SW1H 0ET  
Tel: 020 7215 5000

If you require this publication in an alternative format, email [enquiries@bis.gsi.gov.uk](mailto:enquiries@bis.gsi.gov.uk), or call 020 7215 5000.

**URN 11/1086**