

**SELF ASSESSMENT AS A TOOL TO
MEASURE THE ECONOMIC
IMPACT OF BERR POLICIES - A
BEST PRACTICE GUIDE**

Summary

SEPTEMBER 2009

On 5 June 2009, following completion of this report, the Department for Business, Enterprise and Regulatory Reform (BERR) and the Department for Innovation, Universities and Skills (DIUS) merged to form the Department for Business, Innovation and Skills (BIS).

Self assessment as a tool to measure the economic impact of BERR policies – a best practice guide

Summary

Alice McGee, Debbie Collins and Robin Legard

Prepared for the Department for Business, Enterprise &
Regulatory Reform

P2619



Introduction

This summary is intended to provide a set of ‘best practice’ guidelines for the design and conduct of business surveys, specifically the measurement of economic impact although the general principles apply to **all** surveys. These guidelines are based on findings from a research project that evaluated several BERR surveys, details of which are contained in a number of reports.^{1,2,3} A summary ‘checklist’ of issues to consider when commissioning, contracting and setting up surveys is contained as an addendum to this summary. In addition, this summary also includes our recommendations for further methodological work to explore ways of improving the measurement of economic impact.

A brief background to the research project

The Department for Business, Enterprise & Regulatory Reform (BERR) commissions research that is designed to evaluate the impact of its interventions on improving the productivity and competitiveness of UK businesses. This research aims to determine whether the intervention had the desired impact, or whether this impact would have happened anyway, irrespective of the intervention. Such research questions can be difficult to answer. BERR commissioned the Questionnaire Development and Testing (QDT) Hub at NatCen to develop and test a survey methodology for measuring the impact of its interventions that would overcome these difficulties.

Research aims and objectives

Mode

This research project aimed to understand the difficulties associated with collecting economic impact data, through the use of telephone surveys. Telephone interviewing is the mode most frequently used by BERR to elicit companies’ self-reports on the impact of the support received.

Measuring the counterfactual

A second area of interest focused on measuring the counterfactual, i.e. ‘what would have happened had the policy or programme not been put in place’. This is an important task for BERR in assessing the success of its interventions. However the measurement of the counterfactual is widely agreed to be challenging and the difficulties surrounding its construction require examination.

Research questions

It is widely recognised that experimental methods are valuable in assessing economic impact, specifically the counterfactual. However for economic and practical reasons such methods

¹ McGee A, Collins D, Legard R (2009) Self Assessment as a Tool to Measure *the Economic Impact of BERR policies* – a best practice guide.

² McGee A, Andrews F, Legard R, Collins D (2009) Self Assessment as a Tool to Measure *the Economic Impact of BERR policies: findings from round one cognitive interviews*.

³ McGee A, Andrews F, Legard R (2009) Self Assessment as a Tool to Measure *the Economic Impact of BERR policies: findings from round two cognitive interviews*.

have not tended to be used within BERR to date. Instead BERR has largely viewed beneficiaries as being able to provide evidence on the effectiveness of its interventions and adopted a self-assessment approach to collecting information on economic change and measuring the counterfactual. It is recognised that this methodology is not ideal. A number of questions emerge regarding the reliance on beneficiaries providing evidence of the impact of support products, which were the focus of this research study. The main research questions that this research project sought to address were:

- Can we rely on what ‘beneficiaries’ say about the intervention?
- Are they able to provide us with an accurate description of the economic impact of the support received?
- If so, which would be the best way of obtaining this information from them?
- How can we best use that information to assess the success of the intervention?

This research was concerned with exploring these issues and proposing a new methodology that would improve the validity⁴ and reliability⁵ of economic impact data. It concentrated on one survey in particular – the Business Support Cross Product Monitoring (or “Beneficiaries”) Survey⁶. This is a survey of businesses that use various BERR interventions and includes questions about how these interventions have impacted on the business financially.

Methodology

This research involved a multi-method approach. First, three varying surveys were reviewed and critically assessed using a set of best practice criteria designed for this purpose. The majority of these criteria are fairly general and could be applied to any survey, however there are several that relate specifically to the measurement of economic impact. All the criteria can be found at the end of this summary, in the appendix⁷. These criteria were designed with particular reference to three main sources of guidance.

- ***The Magenta Book, Guidance Notes for Policy Evaluation and Analysis*** – this is a useful guide, produced by the Cabinet Office, for government researchers on the issues to consider in designing surveys, particularly for policy evaluation⁸.
- A set of guidelines, produced by the ***Office for National Statistics*** (ONS) which constitute survey quality and the criteria used for assessing it⁹. It takes as its basis the six dimensions for assessment of survey quality compiled by the European Statistical System (ESS), namely: *Relevance; Accuracy; Timeliness and punctuality; Accessibility and clarity; Comparability and Coherence*¹⁰.
- ***Sources of error in surveys*** drawing on work by Groves et al (2004)¹¹.

Second, questions asked on the Beneficiaries Survey (wave 5) concerned with collecting information on economic change and measuring the counterfactual were subject to cognitive

⁴ Validity refers to the extent of matching, congruence, or 'goodness of fit' between the questions asked and the concept it is purported to measure. i.e. are the questions measuring what was intended?

⁵ Reliability is concerned with questions of stability and consistency (repeatability).

⁶ For a summary of the Beneficiaries Survey please see Appendix A of McGee et al (2009) Self Assessment as a Tool to Measure *the Economic Impact of BERR policies* – a best practice guide.

⁷ For more detail about the design of the best practice criterion please see Chapter 2 McGee et al (2009) op cit.

⁸ (http://www.policyhub.gov.uk/magenta_book/)

⁹ These guidelines can be viewed at <http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=13578>

¹⁰ For more information about these guidelines please see section 2.2.1, McGee et al (2009) op cit.

¹¹ For more information about sources of error please see section 2.2.2, McGee et al (2009) op cit.

testing. The findings from this first round of testing were used to make revisions to these questions, which were then retested¹². Finally, taking on board the findings from both these stages of work, a final set of recommendations were made on the methodology that should be used¹³.

¹² Further details of the cognitive testing stage are contained in the round one and two reports, refer to footnotes 2 and 3 for references.

¹³ For more information on the full set of recommendations please see Chapter 5, McGee et al (2009) op cit.

Recommendations

This summary contains our recommendations, or key messages for those considering conducting surveys to measure economic impact. These fall into three broad categories: (1) specific guidelines for conducting economic impact surveys (each of these arising directly from this research project); (2) further issues to consider when setting up such a survey; and (3) our suggestions for further research into ways to measure economic impact raised by this study.

1 Guidelines for conducting economic impact surveys

This section contains specific recommendations for conducting surveys collecting economic impact data. These guidelines arose from this research project.

(a) Use the most appropriate mode

Evidence from this study suggests that telephone (and specifically Computer Assisted Telephone Interviewing (CATI)) is overall an appropriate mode, well-suited to business surveys particularly for practical and economic considerations. It can collect data of a relatively high quality while achieving reasonable response rates. Telephone surveys are seen as being less intrusive by businesses in comparison to face-to-face surveys and they are less costly to administer. Response rates to a telephone survey are likely to be significantly higher than for a self-completion survey.

However, while telephone interviewing is a mode that appears to be well suited to business surveys, we recommend that mode should be an important methodological consideration for each survey of impact and that an effort should be made to find the mode that is most appropriate to study the research question in hand. For example we recommend the consideration of the use of other modes, particularly as part of a web based **mixed mode approach**¹⁴, to help maintain response rates and improve data quality. Adopting such a strategy might boost (or help prop up) response rates. A second option is to adopt a **mixed methods approach**¹⁵ which would involve providing respondents with a self-completion form for fairly straightforward questions on general business finances (e.g. annual turnover and staff costs) and following up with a telephone interview that would collect more complex information on the intervention and its impact. Another mixed methods approach may be using a combination of both telephone and face to face interviews alongside other types of administrative data such as company accounts.

(b) Frame questions carefully

Use “bite size chunks” within question wording

Question formulation must be given careful consideration to enable meaningful comparison. Questions should be designed so that answers can be given in “bite size chunks”. Long,

¹⁴ Where a mixture of different modes is used (e.g. some respondents complete an online questionnaire and others are interviewed over the telephone).

¹⁵ Where a combination of methods is used (e.g. interviewers administer a face-to-face interview and then leave a self-completion questionnaire for the respondent to fill in and send back).

complex questions, that incorporate multiple terms, are likely to be counter-productive, encouraging the respondent to make up answers rather than give them careful thought.

Use meaningful language

It is important to ensure terminology is meaningful and familiar to all respondent groups thus enabling them to provide as reliable data as possible. In terms of specific examples we recommend replacing the term 'gains' with 'returns', where appropriate. It is anticipated that this will be an improvement to questions using this terminology as the current wording ('gains') may mean answers are biased towards positive responses.

Reference frames and units of observation

Use clear reference frames and units of observation

Reference frames and units of observation must be clearly defined to ensure respondents are using a consistent reference frame in providing their data. Examples are time periods for retrospective questions and units of observation such as 'business' or 'workplace'. Both rounds of cognitive testing showed that respondents thought about different 'units' when answering the test questions because the unit had not been clearly defined to respondents¹⁶.

Standardise responses to a common unit of observation

As well as defining units of observation clearly, where possible respondents should be able to provide information in the format they find easiest. CATI can help with this as questions and response options can be tailored or 'customised' to respondent circumstances. Taking multi-site businesses as an example we recommend using a strategy to **standardise responses to a common unit of observation** (i.e. a workplace or business) rather than asking respondents to restrict their thinking to one or the other. In general, where respondents worked at workplaces or establishments that were part of larger businesses, they would answer the general questions about the business's finances (e.g. annual turnover and staff costs) thinking about the business as a whole and the more specific questions regarding the intervention thinking only about their workplace (the place where they worked). Therefore we recommend customising the wording of questions to refer to 'business' or 'workplace', depending on the respondent's preference and devising a strategy that makes it possible, at the analysis stage, to aggregate these responses to a common observation unit.

In doing this we recommend the inclusion of a set of questions seeking information on the number of workplaces and employees at the outset of the questionnaire to establish the link between workplace and (entire) business. We also recommend the inclusion of 'check' questions throughout the questionnaire to determine whether respondents are thinking of the workplace or the business as a whole when providing financial information. This recommended strategy is based on the current questions used as part of the Workplace Employee Relations Survey (WERS)¹⁷.

¹⁶ For more information about these guidelines please see section 4.1.5 of McGee et al (2009) op cit.

¹⁷ Forth, J. and McNabb, R. (2007) 'WERS 2004 Information and Advice Service Technical Paper No. 1 Innovations' in 'WERS 2004: The Collection of Objective Data on Workplace Performance'. This is available from the NIESR website. <http://www.wers2004.info/technicalpapers.php>

(c) Order questions carefully

Respondents must receive the right encouragement in order to provide accurate responses or even to participate at all. Question ordering plays a role within the accuracy of answers respondents give. We recommend ordering the questions so that factual information is requested initially before asking for an estimate of economic impact. Structuring the questions in this way means respondents will not be asked to give estimates out of the blue.

(d) Target the most appropriate respondent(s)

Use a screening exercise

Although telephone interviewing is felt to be a mode well suited to business surveys it is not necessarily the best way to find the right respondent(s). Depending on the content of the questionnaire and the target population, it is possible that the required survey information may be held by more than one person within a business, particularly for large, multi-site businesses. We recommend the use of a screening exercise to establish the right respondent(s) to answer these detailed questions. This is where a multi-method approach (see section 1(a) above) may be appropriate; face to face interviews could be used within large businesses to make it easier for the interviewer to pin down the right respondent for each specific area of interest within the questionnaire.

Train interviewers

We also recommend spending time, prior to the survey, training interviewers to be able to target the most appropriate individual to answer the questions. Real-life scenarios based on case studies could be useful for this process. Additionally, it may be necessary for the interviewer to make appointments with a number of different people in the organisation to obtain all the required information.

(e) Allow respondents to gather the necessary information outside the interview

Respondents must have the necessary information available to them to answer the survey questions. Respondents need to: (a) know or be able to find out the information being sought (e.g. turnover or exports); and (b) be given the opportunity to obtain the relevant figures after the interview. We recommend that the survey includes a component which enables the respondent to provide further information after the main interview by telephone or post.

(f) Build in sufficient time for an extensive development and testing phase

The developmental stage of a survey is crucial in designing a good instrument, which can obtain high quality economic impact data from respondents. Effective questionnaire design can cross check or verify data given allowing reliable calculations and estimates at the analysis stage. Thus adequate time should be built in to establish: (a) the type of information respondents are likely to have at their disposal; (b) their willingness to divulge it; and (c) their ability to make meaningful estimates about economic impact or change. This information should shape what information is collected, to what level of detail and in what way.

Therefore we recommend that an extensive development and testing phase is conducted. We recommend that this consist of **exploratory interviews** (to help scope out the content of a

questionnaire and identify appropriate terminology) and **cognitive testing** (to help iron out problems well in advance of mainstage fieldwork).

(g) The 'counterfactual' can only be considered an estimate

As the literature shows, and our evidence supports, the counterfactual (what would have happened had the business not received the intervention) is extremely difficult and complex to measure. More broadly, the problem in identifying the impact of an intervention arises from the fact that participation in programmes is not random. Businesses will only take up interventions at 'non-random' times of their lives (e.g. when they are performing very badly or very well) and thus it is difficult to compare these businesses with others that have not received the intervention. It is possible to adequately control for some but not all of the differences between participants and non-participants that might also affect economic outcomes¹⁸.

In terms of whether beneficiaries are able to provide accurate descriptions of economic impact, our review showed there to be a general awareness among respondents about whether or not there had been an effect on business performance (turnover, employment etc) as a direct result of the intervention. However, it was often difficult for respondents to take this further and **quantify** this information in terms of economic impact during the interview. Therefore it is important to be aware of the limitations of this self-reported impact data.

Our review demonstrated that respondents' self-reported accounts of the counterfactual **must** be considered estimates at best. It is important to spend time within the questionnaire, working through the issues logically with the respondent ensuring estimates are informed in that they are based on evidence or experience, rather than guessed at during the interview. These measures should help the respondent to construct an estimate of the counterfactual with some degree of accuracy.

2 Further issues to consider when setting up an economic impact survey

This section outlines further issues to consider or things to be aware of during the planning stages of putting together a survey to measure economic impact. These issues became apparent during our research project as being additional points to consider early on in the process rather than forming specific guidelines that should be followed.

(a) Consider making the survey 'mandatory'

We recommend the consideration of making participating in the 'evaluative' survey a condition of applying for an intervention. This, in our view would boost response to the survey, building on the social exchange theory¹⁹. If respondents see a return to themselves (i.e. the possibility of their application being successful) they will be more motivated to take part in a survey. In addition, for those unsuccessful applicants it will make it easier, potentially, for interviewers to explain why the applicant has been contacted.

¹⁸ Further detail about the counterfactual, specifically regarding 'treated' and 'control' groups can be found in section 4.2.5 of McGee et al (2009) op cit.

¹⁹ For more information about these guidelines please see section 4.1.1 of McGee et al (2009) op cit.

Non-beneficiaries

If the survey is not to be made a mandatory requirement of applying for an intervention it would be important to take non-beneficiaries in account early on, where it is intended that they form part of the sample (by 'non-beneficiaries' we refer to those businesses that applied for an intervention but were unsuccessful). A higher refusal rate is likely among these non-beneficiaries so the sampling strategy would need to take this into account.

(b) Consider using show cards

We recommend the consideration of the possibility of sending out show cards (cards containing pre-specified answer options for particular questions) in advance of the interview. Cognitive respondents were found to struggle to retain all the answer options in their head when they were read aloud. This was found to be particularly difficult when formulating an answer and fitting it into the one of the answer categories was not straightforward. A visual stimulus may be advantageous in assisting with this process, by minimising the response burden of having to remember the answer options.

(c) Consider the use of quantitative and qualitative 'measures'

We recommend the consideration of using a mixture of both quantitative and qualitative 'measures' (or 'objective' and 'subjective' reports) to explore economic impact. The existing literature demonstrates, and our findings from the cognitive testing support that respondents find it difficult to quantify the level of impact an intervention had. Therefore qualitative accounts of what has taken place would also be helpful in fully understanding the effect of the intervention.

(d) Consider building in other components to complement the main survey

Even when using a high quality survey instrument there are limitations to what beneficiaries can provide through self-reported data during a survey interview, when asked to put figures on (or quantify) perceived benefits resulting from an intervention. Further research components, introduced to complement the main survey data, can be invaluable, adding further dimensions, and thus validity, to the research. Therefore we recommend the consideration of such components, these being: (a) the inclusion of a control group to help measure the counterfactual; (b) conducting case study interviews for additional evidence purposes; (c) the use of other existing data sources to cross reference with and enhance the survey estimates; and (d) the use of a mixed methods approach (for instance using a combination of face to face and telephone interviews alongside company accounts and other administrative data).

Existing data sources

There are two main types of existing data source that could be harnessed. These are:

- (1) registers of all businesses across Britain that include economic data that could be compared with our survey data.
- (2) data originating from the original application data for a business included in the survey sample.

Further details on how these data sources could be used is contained in section 3(g).

3 Suggestions for wider research

This final section contains our recommendations for wider research. These are intended as broader areas for BERR to consider in commissioning further exploratory research, rather than issues that could be addressed within one specific project.

(a) Further explore Beneficiaries questions and survey data

We feel there is potential in further testing and validation of the revised questions proposed for the Beneficiaries survey and secondary analysis of existing survey data.

Further validate proposed new questions (Beneficiaries survey)

We feel there would be value in the Department exploring ways to further validate the Beneficiaries questions, tested and revised as part of this project. We have conducted two rounds of cognitive testing and made revisions in light of the findings. However, the final recommended questions have not been subjected to further cognitive testing or to field-testing. While we feel the alterations we have made are likely to constitute improvements it would be beneficial to further test these questions to find whether this hypothesis is true. Cognitive testing would be of value in this regard as would some form of split panel experiment, which would enable differences in data quality obtained to be quantified.

Explore the nature of inconsistencies between financial data and self-assessed economic impact

We also recommend that further secondary analysis of the Beneficiaries Survey data should be undertaken to explore the nature of inconsistencies between financial data and self-assessed economic impact. For example, such analyses could investigate whether certain kinds of businesses, such as those that are relatively new or in particular sectors, are more likely to provide inconsistent data.

(b) Explore whether sending advance information could improve data quality

Evidence from this research project highlighted that questions asked over the telephone without any prior warning could encourage respondents to guess the answer rather than providing accurate information. A well designed letter, sent in advance of the interviewer making contact, can be used to (a) encourage participation; and (b) alert respondents to the types of information that will be asked for/questions to be asked. However, such a letter can act as an incentive or a deterrent to participation depending on how accessible the information. We recommend this to be an area requiring further exploration as this approach could potentially improve data quality.

(c) Commission research into the overall timing of the survey

The early stages of this project flagged up the issue of the timing of the survey (i.e. how long after the intervention had been received should firms be contacted). As some surveys cover questions on (a) the company's rationale for application and (b) the economic impact of the intervention, there is clearly a balance to be struck in terms of finding when the survey should be conducted (i.e. how soon following participation in the intervention). In order to fully assess the best time to ask these questions regarding impact of an intervention we recommend conducting secondary analysis on existing data, following this up with an experiment to compare the sorts of information gathered, looking at following up at different points in time (e.g. 3 months, 6 months or 1 year following the end of receipt of intervention, for example).

(d) Further explore the use of a longitudinal design

Taking the issue of timing further still we recommend the Department consider the possibility of using a longitudinal rather than a cross-sectional approach. This would help address the problem of simply having a snapshot of the business following the intervention and instead gain a more accurate picture over a period of time. It has been shown that situations can change greatly over a relatively short period of time and we are not currently capturing this. Wide variability exists within the respondent group and further follow up interviews would help better explore this variability.

(e) Explore the appropriateness of financial measures

A further issue that requires consideration is the appropriateness of the financial measures of impact, which are central to the current impact assessment strategy. If the Department were willing and able to utilise longitudinal administrative data on firms, linking this to programme participation, then it would be possible to supplement this approach with analyses of firm survival and growth. These are two alternative measures of performance, both highly correlated with profitability, which are commonly investigated in the industrial and labour economics literature.

(f) Conduct further qualitative work to further explore the counterfactual

We recommend the Department look into commissioning more qualitative work in this area. For example depth interviews could be conducted with respondents to validate their answers, exploring their circumstances in more depth to understand whether accurate answers had been given to the survey they took part in. More generally it would be beneficial to collect more qualitative data on respondents' motivations for applying for the intervention and what they expect to get out of it long term (employment growth, survival etc). These findings could be compared with the survey findings, adding to the explanatory power of the study.

(g) Explore alternatives to self-assessed impact data

We feel that there is sufficient evidence to indicate that sole reliance on data obtained from subjective self-assessments, currently collected in many BERR surveys, can hinder the Department's efforts to assess the economic impacts of its policies and programmes. We recommend the Department explores alternative data sources and outline three possible options to consider below:

1. Investigate the practicality of alternative retrospective assessments along the lines proposed by Manski (2004)²⁰.
2. Consider the possibility of before-after sampling using accounting information on the performance of participants pre- and post-participation. This would provide some objective measure of what 'really' happened to the firm, which at the very least would provide useful validation of the success of the current self-assessment methodology.
3. Seriously consider alternative strategies for identifying the causal effect of these schemes, including before-after surveys with 'like' non-participants.

²⁰ Manski, C. (2004) 'Measuring Expectations'. *Econometrica*, 72,5, 1329-1376. For more information please see section 4.2.4 of McGee et al (2009) op cit.

Clearly the costs and benefits of spending more money and effort on programme evaluation need to be considered. This will depend, to some degree, on the extent to which these additional data sources are likely to add to the survey data in terms of identifying impacts which would otherwise be missed or identifying where the impacts reported from existing survey data may be overestimated. However we believe the Department should be looking to a strategy that places greater emphasis on utilising the longitudinal administrative data they have on participant and 'like' non-participant firms, which can be linked to firm performance data. Survey data provide information about process, for example how things are or are not bedding in, and in further understanding of how firms orient themselves towards government support. Combining administrative and survey data would provide a richer and more accurate picture of the impact on BERR policies on businesses.

(h) Explore the use of existing data sources in constructing the counterfactual

There are data sources available that might form the basis for a quasi-experimental analysis of intervention effects. Two of these are identified below:

Inter-Departmental Business Register and Annual Business Inquiry

The Inter-Departmental Business Register²¹ (IDBR) offers a sampling frame that is the universe of all firms in Britain registered for VAT or PAYE purposes. It is possible to use unique identifiers in this register to link businesses to other data sets, including programme evaluation databases, and thus this could form the basis for studies comparing participants with non-participants. These studies could use existing data sources, such as the Annual Business Inquiry²² as the basis for economic data. Alternatively, the IDBR could be used to sample organisations for a survey-based evaluation. BERR may therefore wish to consider using these data in one of two ways.

1. Data for businesses within the sample (i.e. that received the intervention) could be taken and compared with answers to survey questions; this approach could then be used to assess the success of the survey measures.
2. Data for potentially similar businesses to those in the sample could be taken and used to form a 'control' group to compare with similar administrative data for businesses within the sample (i.e. that received the intervention).

Intervention application data

Additionally we recommend the Department explore the option of harnessing the original application data for a particular business. When the decision was made to award the intervention the 'gatekeeper' must have used some kind of scoring system to determine or assess eligibility (e.g. profiling, score assessment, selection process) and made some subjective assumptions regarding how well a company would perform having received the intervention. It would be useful to compare this information with the survey outcomes.

(i) Consider further research into how random assignment might alleviate bias

Finally, we would like to raise the possibility of using random allocation methods to assign firms to BERR programmes. Random assignment is the 'gold standard' in terms of identifying causal impacts of programme participation. The reason for this is that random assignment, if undertaken properly, ensures that participation in the programme is not associated with other

²¹ <http://www.statistics.gov.uk/idbr/idbr.asp>

factors about the participant firm which are unobservable to the evaluator but may be correlated both with participation and the outcome of interest. That is to say, random assignment ensures that impact estimates are not confounded by biases associated with unobservable differences between participants and non-participants. We recommend the Department consider further exploration of this option. It is worth noting that random assignment has recently been adopted by DWP as the preferred method in evaluating labour market programmes.

²² <http://www.statistics.gov.uk/abi/>

Summary appendix: Evaluative criteria

A set of general criteria were developed to evaluate the three BERR 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
1. Research aims	
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 measure . 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.
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 steering group to assist in shaping the research and ensuring it remains on the right track. A feasibility test or small scale pilot test may be useful in helping ensure the survey design is robust.
2. Definitions	
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.

Criteria	Description
3. Target population	
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).
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.
4. Data collecting mode	
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.
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.

Criteria	Description
5. Questionnaire design	
a) Questionnaire content established	The content and priorities of the questionnaire should be decided after re-viewing the survey aims. A period of developmental work 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.
b) Most appropriate respondent(s) identified	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.
c) Questions designed that respondents understand, are willing and able to answer	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.
d) Questionnaire fully tested	The questionnaire must be subjected to rigorous testing 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 ²³ . 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.

²³ Collins, D (2003) 'Pretesting survey instruments: An overview of cognitive methods' Quality of Life Research 12:229-238.

Criteria	Description
6. Data collection	
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.
7. Data processing	
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.
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.
8. Data analysis	
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 ²⁴ .

²⁴ 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
9. Outputs and dissemination	
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.
10. Management issues	
a) Successful commissioning process	The commissioning process should aim to follow the Social Research Association guidelines (pages 6 to 8) ²⁵ 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.
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 the findings will be used by policy customers 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.

²⁵ The SRA guidelines can be found at http://www.the-sra.org.uk/commissioning_sr.htm

Specific criteria evaluating the measurement of economic impact

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 actual change the policy or initiative had on the company, respondents must be fully equipped to be able to consider the counterfactual : 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.

