

# DECC smart meters evaluation and consumer benefits measurement workshop



Prepared by: Gary Raw  
Regional Director

Checked by: David Ross  
Regional Director

Approved by: David Ross  
Regional Director

The initial DECC smart meter evaluation workshop

Rev No	Comments	Checked by	Approved by	Date
1	First draft to client	David Ross	David Ross	21/09/11
2	Revisions following client comments	David Ross	David Ross	23/09/11
3	Revisions following circulation to participants	David Ross	David Ross	13/10/11

AECOM House, 63-77 Victoria Street, St Albans, Hertfordshire, AL1 3ER  
Telephone: 01727 535000 Website: <http://www.aecom.com>

Job No 60221038

Date Created: September 2011

This document has been prepared by AECOM Limited for the sole use of our client (the "Client") and in accordance with generally accepted consultancy principles, the budget for fees and the terms of reference agreed between AECOM Limited and the Client. Any information provided by third parties and referred to herein has not been checked or verified by AECOM Limited, unless otherwise expressly stated in the document. No third party may rely upon this document without the prior and express written agreement of AECOM Limited.

c:\users\gary\documents\documents\gr people solutions\decc\sm evaluation\workshop report\sm evaluation - workshop report - v5.1 gjr.docx

## Acknowledgements

The author is grateful to David Shipworth (UCL Energy Institute), Michael Harrison (DECC) and David Ross (AECOM) for discussion during the preparation of the workshop and this report. In addition, David Shipworth, David Ross and Demelza Birch facilitated table discussions and feedback sessions, without which the workshop would not have been possible. Thanks also to Maxine Frerk (DECC) for opening the workshop, Siobhan Campbell (DECC) for presenting on organisation of the evaluation and Michael Harrison for making notes throughout. Finally, the workshop was possible because participants gave their time voluntarily and particular thanks are due to all of them.

## Table of contents

Summary.....	1
1 Introduction .....	5
2 The workshop process.....	6
3 Priorities for evaluation criteria.....	9
4 Roll-out activities .....	17
5 Evaluation good practice .....	18
6 Organisation of the evaluation .....	23
7 Remaining issues.....	24
8 Conclusions.....	25
Annex A. Supplementary Documents .....	31
Annex B. Criteria for evaluation.....	32
Annex C. Detailed findings on evaluation criteria .....	33

Capabilities on project:  
Building Engineering

# Summary

## Introduction

This report is the final deliverable from AECOM's scoping study for energy consumption change measurement in the smart meter roll-out. It documents the content of, and main messages from, a workshop held on 8 September 2011. The aim was to start the process of DECC working with suppliers and other external experts on the detailed arrangements for the roll-out evaluation and benefits measurement. The focus is on how to evaluate benefits related to energy consumption change during the Foundation stage. The key points emerging from the workshop are summarised here but the main text includes details that may be relevant to particular decisions in future.

## The purpose of the evaluation

Evaluation of the smart meter roll-out is important:

- to give greater certainty over the best approach and the likely impacts (especially on householder behaviour);
- to provide feedback during the roll-out, to improve the roll-out process and provide early warning of risks;
- for suppliers to maximise the benefit to themselves, customers and UK commitments to reduce energy demand;
- to give the public (and NAO) confidence that the roll-out is of value and being conducted fairly and efficiently by Government and the supply industry (thus maintaining credibility of the Smart Meter Programme);
- to benefit from overlaps with other policies and programmes; and
- because there is a commitment to do it.

## Evaluation criteria

There are many possible evaluation criteria (i.e. ways of judging the success of the roll-out). The workshop rated a set of 36 possible criteria on their *importance* and the *quality* of measurement (precision, accuracy or confidence) that would be appropriate for each. Five general points can be made from the ratings.

- The rank order of ratings provides a starting point for deciding which criteria to use and the quality of measurement that should be planned.
- In general, participants thought that if something was worth measuring, it was worth measuring well.
- Ratings of importance were highest for the following criteria relating to the Foundation stage roll-out as a whole:
  - savings on domestic energy bills;
  - reduced overall demand;
  - customer convenience/satisfaction;
  - changes in consumer behaviour;
  - quality of consumer trials;
  - quality of smart metering installation, services and support;
  - quality of energy supply offerings;
  - quality of information, feedback and support available to help consumers choose the most beneficial products and services;
  - uptake of new tariffs and other smart-enabled products and services;
  - data security and privacy.
- The first five of these also attracted the highest importance ratings in relation to evaluation of specific products and services, as distinct from the roll-out as a whole. In general, criteria relating to specific products, services and consumer engagement strategies had intermediate importance ratings.
- Suppliers tended to give higher ratings than other participants for quality requirement and slightly higher for importance. Criteria may be placed in rank order by these differences, indicating where DECC might need to take the initiative in relation to a particular criterion.

Capabilities on project:  
Building Engineering

The key general points from the discussion and from comments made on individual criteria are as follows.

- Some criteria may be more relevant to long-term evaluation than to the Foundation (e.g. because benefits would take time to appear, or there would be too few smart meters to evaluate outcomes, or early evidence could give a misleading impression of what would happen in the full roll-out). Nevertheless, the Foundation stage can be used to prepare for later evaluation by establishing measurement processes, learning about impacts and building up data on the baseline position (accepting that impacts may differ between Foundation and main stages).
- Selection of criteria also needs to take account of other industry stakeholders and the trade-offs between the benefits of carrying out evaluation, the existing sources of data and the costs and difficulty of measurements.
- DECC will need to decide the level at which it wishes to evaluate, what could be taken forward with partnerships (e.g. with academic researchers) and what is left to the industry. If the essential requirement is to evaluate impacts across the total population (in line with the benefits case) and the delivery of outcomes, then the paths/actions that suppliers are taking could be left to them (unless the roll-out was in some sense found to be failing). Nevertheless, in order to be confident of overall impact, a wide range of data may be needed, and specific areas of difficulty identified. DECC's scope also needs to take account of possible bias in suppliers' findings as a result of their desire to understand what works for specific segments and to understand the mechanisms of particular interventions.

## Organisation of the evaluation

Organisation of the evaluation is about how DECC, suppliers and other parties will work together to deliver the evaluation, avoiding duplication of effort and setting and assuring standards to ensure a minimum level of quality without suppressing creativity and innovation by suppliers.

For the suppliers, in the Foundation stage, the priority is gaining access to customers, and getting the basic metering systems working and interoperable. Development and assessment of products and services then follows, with progression into measuring changes in consumption as the programme moves into mass roll-out. In public policy terms, the focus is on whether benefits are being realised. In this context, confidence in the evaluation may be enhanced by linking it to benefits identified in the DECC impact assessment.

In the Foundation stage, there is no obligation on suppliers to do anything in relation to the roll-out. Commercial sensitivity was therefore a recurring theme but, in practice, it will depend on the type of information, the format and time frames in which it is provided and to whom it is provided. In general, suppliers were more comfortable sharing information in confidence with Government (at regional/national level) than with competitors. Suppliers indicated that the onus is on DECC to propose a plan for the data required, who will be granted access and on what basis. This is likely to apply to some extent throughout the roll-out, not just in the Foundation stage.

This is part of a broader message from the workshop, that suppliers are looking to DECC (in consultation with all interested parties) to specify what additional activities (if any) and what data will be expected, and what would give DECC confidence that benefits were being delivered and consistently assessed. This was seen as urgent, to achieve consistency of methods and data and for suppliers to know what is expected of them. Suppliers also expressed a need to "keep it simple" with, as far as possible, a business-as-usual solution at least cost.

Any targets should also be translated into information for customers, to retain their confidence and avoid unnecessary disappointment (e.g. if suppliers cannot meet demand for smart meters, or if customers have an unrealistic expectation of what savings they can achieve).

Two other key activities for DECC were seen as establishing a central data collection point, such as the proposed DataCommsCo (DCC), and delivering common and consistent central messages, to mitigate programme risks that

Capabilities on project:  
Building Engineering

could distort impact and take-up (e.g. concerns over electric and magnetic fields (EMF) or perception that the programme is too expensive).

## Execution of the evaluation

### Evaluation design

External validity is compromised by self-selection, e.g. customers opting in to installation or allowing suppliers access to detailed data. This would largely be overcome by basing evaluation of standard installations on customers selected on an opt-out basis (opt-in may be required for at least some additional smart-enabled products and services). This approach would require a control group and there is no basis for preventing the control group having a smart meter installed if customers request it, so it would need to be large enough to allow for these losses.

An opt-in sample does not give valid findings for people who are choosing the intervention – the important point is not to generalise those findings to people who have not chosen the intervention without first accounting for the likely biases. Opt-in studies can also identify interested customer segments and find out what gets them to ‘sign up’.

The possibility was discussed of creating control groups by installing a smart meter and not telling the household about it (e.g. by presenting the installation as business-as-usual meter replacement). However, this may not be a viable approach, in terms of either consumer protection or compliance with other Programme requirements.

Internal validity is compromised because, as noted above, smart meter installation cannot be restricted and neither can some other important variables (see further comments below, under “Data collection”).

### Sampling and recruitment

Suppliers' strategies will be influenced by practical/commercial factors. In the Foundation stage, there is likely to be a tendency to avoid hard-to-install properties, prepayment customers and consumers with less benefit/engagement (e.g. those with low energy use) but prioritise customers where there is the opportunity to reduce avoidable debt. This would bias results in relation to expectations for the full roll-out; therefore we need to understand whether early mover installations are representative. Two approaches were identified for dealing with this:

- accept that lessons need to be learned from easier installations before moving on to the more difficult ones;
- expect some difficult installations to be made in the early Foundation phase, to ensure that practical issues are fully understood and resolved and differences in impact are identified (this could target all homes in a limited geographical area and may require Government funding support).

An upfront and honest approach to recruitment will generate a better response from consumers – to do otherwise would run the risk of the roll-out being halted because of privacy issues and campaigns.

### Data collection

Two kinds of data collection were envisaged: routine data flows, and collecting data specifically to evaluate the roll-out. Routine data flows were seen as being limited in several ways.

- Evaluation requires good baseline data and this is limited. Where it depends on non-smart meter data (with relatively poor quality), larger numbers of homes and long periods will need to be used. Alternatively, where additional smart-enabled services are being evaluated, smart meter data could be used (but for relatively small numbers of homes in the Foundation stage).
- There were particular concerns over the early (pre-DCC) period. Suppliers are building complex data systems and it will be challenging to get data out of these (if requirements are not specified early enough).

Capabilities on project:  
Building Engineering

- The Annual Quantity database (of consumption) is useful for looking across regions but possibly not for understanding impacts of individual interventions. Also, it will be important to extract these data at the right point in processing, e.g. before/after temperature correction.
- ELEXON's profile sampling meter base may be relevant, providing accurate, representative data.
- The timing of smart meter installation will affect the apparent impact, so data on timing of installations is required.

It will not be feasible to measure everything and there is a trade-off between doing a few things well or many poorly. In addition, suppliers may already have their own trial objectives and data requirements (it could be difficult to add to these) and there could be customer resistance to collecting more data about them than suppliers do currently. However, it will be important to have sufficient data to detect and quantify effects.

Energy use is dependent on many factors that are not the main focus of suppliers' interests. Also, householders may change supplier, undertake installations or change their behaviour in ways that will create 'noise' in the data if they are not measured or observed. Energy suppliers lack the wider data (including timing of changes) that would be needed to interpret such extraneous changes. This implies a need to create new data linkages and/or facilitate dedicated data collection as part of the roll-out generally or in specific trials. The implied need for a wide range of measurements and observations suggests that only small samples will be feasible with the full range of variables. Close matching of samples would mean that much smaller sample sizes would be needed in order to achieve statistically significant effects.

A further source of variation that needs to be controlled for is consistency of message in customer contacts. Suppliers saw this as difficult to manage because contacts may go to different points within a supplier organisation (e.g. general call centres, call centres dedicated to smart meters, or the installers), resulting in different messages.

### **Customer feedback**

This is a specific aspect of data collection and the general points above also apply here.

There is potential value in creating a short harmonised questionnaire to use in all data collection exercises. Questions would always need to be asked in the same sequence (before other, non-standard questions). It also matters who administers a questionnaire, so there would need to be a standard type of interviewer (e.g. supplier, Government, University) – this was seen as important in relation to the potential for bias in customer responses (e.g. over-estimating or over-reporting changes in behaviour) or willingness to provide personal household data at all. A market research organisation, independent of the industry and Government, may get the most accurate perspective.

This raises issues over linking datasets (e.g. because of data protection requirements) so that either the energy supplier has feedback data or the other organisation has consumption data, at household level. Data privacy was also seen more generally as a barrier, particularly if customers have the right to opt out of data linking.

### **Data analysis**

An important general point concerned who should carry out the analysis. This related to the conduct of the analysis but, more importantly, to independence and trust in the findings. Data collection/evaluation by an independent body or Government may be better received by consumers. Similarly, it is important to be clear about the audience for the analysis (e.g. suppliers, DECC, Ofgem, the public) and what data/analysis can be aggregated for public consumption or used internally (by suppliers or Government) for refinement of the Programme.

A particular challenge for the analysis will be multiple policies impacting on the same outcomes. The analysis needs to take account of the timing of other interventions (e.g. Green Deal installations). It will also need to take into account varying approaches to the Foundation stage (actions trialled in the Foundation stage could create bias later) and to monitor additional third party interventions.

Capabilities on project:  
Building Engineering

# 1 Introduction

This report is the final deliverable from AECOM's scoping study for energy consumption change measurement in the smart meter roll-out. It documents the content of, and main messages from, a workshop held on 8 September 2011. The workshop participants were a mix of suppliers, DECC officials and invited experts. The aim was to start the process of DECC working with suppliers and other external experts on the detailed arrangements for roll-out evaluation and benefits measurement. The focus is on how to evaluate benefits related to energy consumption change during the Foundation stage. Specifically, the workshop aimed to:

- develop a common understanding of the why, what and how of the evaluation;
- discuss key requirements;
- identify potential problems and suggest possible solutions.

The principal application of the report will be to inform DECC decisions on how to take forward the evaluation of the smart meter roll-out, including any pilot studies, with particular reference to consumer energy savings in the early (Foundation) stage. The Government's impact assessment necessarily identifies key issues in relation to the anticipated benefits of the smart meter roll-out. This gives a clear signal in relation to evaluation priorities but does not completely define the evaluation content or focus.

Application will depend on the general terms agreed for the evaluation, such as whether:

- the evaluation uses only routine monitoring data or also experimental trials with subsets of homes;
- any trials are entirely within the remit of the suppliers conducting them or there is central coordination of trials;
- evaluation of the smart meter roll-out is conducted in isolation or coordinated with the design and/or evaluation of other programmes (e.g. Green Deal).

This will affect, for example:

- the selection and combination of interventions aimed at reducing energy demand through behavioural change;
- the potential for targeting particular types of household or behaviour;
- whether households are knowingly recruited to the evaluation or merely monitored;
- whether control groups are preselected or defined retrospectively using monitoring data.

The next chapter describes the processes used in the workshop and notes who the participants were. The following chapters summarise the learning, following the programme of the workshop itself.

- Chapter 3: priorities for evaluation criteria.
- Chapter 4: activities that might be included in the roll-out.
- Chapter 5: evaluation good practice.
- Chapter 6: the organisation of the roll-out evaluation as a whole.
- Chapter 7: identification of suppliers' remaining issues.

Chapter 8 draws general conclusions from across all the sessions.

Annex A lists the separate Supplementary Documents that support this report and Annex B lists the evaluation criteria considered during the workshop. Annex C provides more detailed findings on the evaluation criteria, to supplement Chapter 3.

Capabilities on project:  
Building Engineering

## 2 The workshop process

### 2.1 Invitations and attendees

DECC, in consultation with AECOM, selected invitees to represent a mix of suppliers, Government officials and external experts. Letters of invitation<sup>1</sup> were issued by email from DECC on 15 August 2011. A two-page background paper<sup>2</sup>, prepared by AECOM in consultation with DECC, was sent to interested parties on 30 August 2011. DECC also followed up these communications to encourage attendance.

The final list of attendees (in addition to Gary Raw and David Ross of AECOM) was as follows.

<i>DECC</i>	<i>Energy supply companies</i>	<i>Other experts</i>
Demelza Birch	David Anderson (British Gas)	David Jones (ELEXON)
Siobhan Campbell	Paul Clark (SSE)	Rob Metcalfe (University of Oxford)
Maxine Frerk	Hannah Darby (First Utility)	Gill Owen (Sustainability First)
Michael Harrison	Steve James (E.ON)	Hazel Preston-Barnes (PassivSystems)
Kate Hughes	Liz Kenny (NPower)	David Shipworth (UCL Energy Institute)
	Ashley Pocock (EDF Energy)	Michelle Shipworth (UCL Energy Institute)
	Graham Smith (Scottish Power)	Jason Stevens (ERA / Engage Consulting)

DECC's specification of a workshop of about 18 attendees was therefore realised.

Parts of the workshop involved exercises in three groups, at separate tables (where participants sat throughout the workshop). For this purpose, attendees were assigned to groups of six or seven, each having a mix of suppliers, officials and other experts. A facilitator was pre-assigned to each table.

### 2.2 Workshop content

The workshop followed a timetable agreed with DECC (see Table 2.1). The details are described in the following chapters and the slides shown during the workshop are in Supplementary Document C. The following paragraphs give an overview of the workshop content, to place each substantive session in context.

**Session 3 (Introduction to the day)**<sup>3</sup> restated the context and aims of the workshop and gave a brief overview of the timetable.

**Session 4 (Why evaluation?)**<sup>4</sup> summarised the key reasons why evaluation should be conducted at all and why it – together with benefits measurement – should be taken seriously. This includes: because there are uncertainties over the best approach and the likely impacts (especially on householder behaviour); to provide feedback to improve the roll-out process; to provide early warning of risks; for suppliers to maximise the benefit to themselves, customers and UK commitments to reduce energy demand; to give the public (and NAO) confidence that the roll-out is of value and being conducted fairly and efficiently by Government and the supply industry (thus maintaining credibility of the Smart Meter Programme); and because there is a commitment to do it.

<sup>1</sup> See Supplementary Document A.

<sup>2</sup> See Supplementary Document B.

<sup>3</sup> Slides 2-5 in Supplementary Document C.

<sup>4</sup> Slides 6-9 in Supplementary Document C.

Capabilities on project:  
Building Engineering

**Table 2.1** Workshop timetable

Session number / title		Session content/aim
1	Arrive, refreshments	
2	Welcome and introductions	Opening words, domestic arrangements and introduction of participants and workshop organisers.
3	Introduction to the day	Run through aims and timetable.
4	Why evaluation?	Explain rationale for evaluation.
5	Evaluation criteria	Explain what is meant by criteria; identify possible criteria; categorise criteria according to their importance and the desirable level of precision, accuracy or confidence.
6	Buffet lunch	
7	Feedback on session 5	Agreement, disagreement and the reasons for them.
8	Roll-out activities	Overview of possible actions, focusing on early trials and roll-out activity.
9	Evaluation good practice	Overview of basics of good practice in evaluation.
10	Practical issues with good practice	To identify likely problems with implementing good practice.
11	Solving problems	Develop ideas for solving problems identified in session 10.
12	Managing standards	How standards will set and assured for evaluation design and implementation.
13	Preparing the way	A chance to clarify where support will most be needed if the evaluation is to be a success.
14	Final comments	
15	Close and refreshments	

**Sessions 5 and 7 (Evaluation criteria)**<sup>5</sup> started with presentation of an overview of possible evaluation criteria, i.e. the basis on which the success of the roll-out might be judged. This was structured around six ways of characterising the criteria (the potential beneficiaries, scope of the evaluation, the roll-out process, the roll-out impact, the source of data, and synergy with other actions). These six factors were used to create 36 criteria for workshop participants to evaluate. Working in table groups, they made judgements on the *importance* of each criterion for evaluation and the quality of measurement (*precision/accuracy/confidence*) that would be appropriate. The aim was to identify shared interests in good evaluation and where there is a divergence of views that will need to be resolved. In plenary feedback session, key points arising from the table discussions were highlighted and discussed further.

**Session 8 (Roll-out activities)**<sup>6</sup> was a presentation and discussion of possible supplier activities in early trials and roll-out activity. This was to provide DECC with insight into the range of contexts for the evaluation.

**Sessions 9-11 (Evaluation good practice)**<sup>7</sup> began with a presentation of some basics of good practice in evaluation. This used headings of design, sampling, recruitment, data collection, customer feedback and data analysis. Participants then individually contributed suggestions as to the possible practical problems with adopting each aspect of good practice. In the following plenary discussion, the potential problems were summarised and – where possible – solutions suggested.

<sup>5</sup> Slides 10-20 in Supplementary Document C.

<sup>6</sup> Slides 21-22 in Supplementary Document C.

<sup>7</sup> Slides 23-38 in Supplementary Document C.

Capabilities on project:  
Building Engineering

**Session 12 (Managing standards)**<sup>8</sup> addressed the overall organisation of the evaluation – how standards will be set and assured to ensure a minimum level of quality without suppressing creativity and innovation. There was a brief presentation and discussion of different approaches to applying standards.

**Session 13 (Preparing the way)** was an opportunity for suppliers to say which aspects of evaluation they were likely to find most difficult and what support they are most likely to need (e.g. what expertise, input, topics or types of person/organisation). The other parties at their table were available to help them to refine, detail and clarify the need for support. This was to assist DECC in understanding suppliers' issues, to prepare the ground for setting out reasonable expectations.

At the end of the workshop, all written records were collected and a transcript was produced as raw material for this report. Participants were also invited to send DECC any additional comments, by 15 September, and to comment on a draft of this report (no such comments were received).

---

<sup>8</sup> Slides 39-42 in Supplementary Document C.

Capabilities on project:  
Building Engineering

## 3 Priorities for evaluation criteria

### 3.1 Method

This part of the workshop addressed the issue of evaluation criteria, i.e. the basis on which the success of the roll-out might be judged. The aim was to identify shared interests in good evaluation and where there is a divergence of views that will need to be resolved.

AECOM presented an overview of possible criteria, structured around six ways of characterising the criteria (the potential beneficiaries, scope of the evaluation, the roll-out process, the roll-out impact, the source of data, and synergy with other actions).<sup>9</sup> Prior to the workshop, these six factors (mainly the scope of the evaluation, the roll-out process and the roll-out impact) were used to create 36 criteria for workshop participants to evaluate (see Annex B). Working at the three tables (Tables A, B and C), participants were presented with the criteria in three groups (Group 1 was criteria 1-11, Group 2 was 12-22 and Group 3 was 23-36) and they evaluated each group before moving on to the next. Each table received the criteria Groups in a different order (Table A: 1-2-3, Table B: 3-1-2 and Table C: 2-3-1).

The evaluation combined judgements on the *importance* of each criterion for roll-out evaluation and the desired quality of measurement (*precision/accuracy/confidence* – “PAC”). In other words, should the criterion be used at all and, if so, should measurements be made with high precision, accuracy and confidence? These judgements related to the Foundation stage. Participants (excluding facilitators) first made their own judgment on each criterion in the group, based on the grid shown below (the numbers are for easy reference to the nine grid boxes). Participants were asked to use only the four corner boxes at this stage. They recorded their votes on sheets as shown in Supplementary Document D and these sheets were collected by AECOM for further analysis.

		Importance		
		Low		High
Accuracy, precision or confidence	High	1	2	3
		4	5	6
	Low	7	8	9

The whole table then discussed the criteria and agreed where they should be placed on a table-sized version of the grid. This was done using sticky notes pre-printed with the criteria. At this point, the intermediate boxes could also be used if necessary. The discussion was recorded on flipcharts at each table.

In plenary feedback session, key points arising from the table discussions were highlighted and discussed further. Overall, the discussion was aimed at understanding the reasons for “votes”, agreement and disagreement, to guide the final selection of criteria.

<sup>9</sup> See slides 10-20 in Supplementary Document C. Participants had a paper copy of the presentation for easy reference during the discussion and they were reminded of this at the start of the presentation.

Capabilities on project:  
Building Engineering

For analysis of evaluations by individual participants, “votes” were converted to ratings for importance and quality (PAC). For importance, boxes 1, 4 and 7 scored zero, 2, 5 and 8 scored a half and 3, 6 and 9 scored one. For PAC, boxes 7, 8 and 9 scored zero, 4, 5 and 6 scored a half and 1, 2 and 3 scored one. Means and standard deviations were calculated for each criterion on each measure (for all participants and separately for suppliers and other participants). No statistical tests were applied to differences because (a) sample numbers are very small and (b) the conclusions relate to the particular group of people who participated in the workshop, rather than treating them as a sample of a population.

## 3.2 Findings

### 3.2.1 Ratings at table group level

The final placement of the 36 criteria by each group is shown in Figure C1 (see Annex C). Refer to Annex B for the criterion definitions (readers may wish to print this Annex separately for easy reference when reading the following text, in which the criterion numbers have generally been used rather than spelling out each criterion in full every time it is mentioned). The following observations may be made on these placements.

- All three tables agreed on placing criteria 6, 19, 20 and 23 in box 9 (high importance, high PAC).<sup>10</sup>
- In addition, all three tables agreed on high importance for criteria 1, 3, 4, 7, 9, 10, 24, 26 and 28.<sup>11</sup>
- All three tables agreed on high PAC for criterion 21.<sup>12</sup>
- All three tables agreed on low importance for criterion 32<sup>13</sup> and low PAC for criterion 35<sup>14</sup>.
- In no case did all three tables agree on intermediate importance for a criterion, and placement in the intermediate column was such that tables differed by more than one column in eight cases (criteria 12, 13, 14, 18, 21, 30, 33 and 36).<sup>15</sup>

---

<sup>10</sup> 6 = Savings on domestic energy bills in the roll-out as a whole. 19 = The quality of consumer trials in the roll-out as a whole. 20 = Reduced overall energy demand in the roll-out as a whole. 23 = Savings on domestic energy bills in relation to specific products and services.

<sup>11</sup> 1 = The quality of smart metering installation, services and support in the roll-out as a whole. 3 = The quality of information, feedback and support available to help consumers choose the most beneficial products and services in the roll-out as a whole. 4 = The quality of data security and privacy in the roll-out as a whole. 7 = Improvements in customer convenience in the roll-out as a whole. 9 = Changes in consumers' energy-related behaviour in the roll-out as a whole. 10 = The uptake of smart-enabled products and services (other than new tariffs) in the roll-out as a whole. 24 = Improvements in customer satisfaction in relation to specific products and services. 26 = Reduced overall energy demand in relation to specific products and services. 28 = The quality of consumer trials in relation to specific products and services.

<sup>12</sup> 21 = Smoother time profile of domestic energy use in the roll-out as a whole.

<sup>13</sup> 32 = Improvements in customer management efficiency (e.g. process simplification) in relation to specific products and services.

<sup>14</sup> 35 = The uptake of smart-enabled products and services (other than new tariffs) in relation to specific consumer engagement approaches.

<sup>15</sup> 12 = Improvements in energy supply efficiency in the roll-out as a whole. 13 = Improvements in energy supply capacity in the roll-out as a whole. 14 = Improvements in customer management efficiency (e.g. process simplification) in the roll-out as a whole. 18 = Improvements in energy generation capacity in the roll-out as a whole. 19 = The quality of consumer trials in the roll-out as a whole. 21 = Smoother time profile of domestic energy use in the roll-out as a whole. 30 = Improvements in energy supply efficiency in relation to specific products and services. 33 = Improvements in industry cost-effectiveness in relation to specific products and services. 36 = The uptake of new tariffs in relation to specific consumer engagement approaches.

Capabilities on project:  
Building Engineering

- All three tables agreed on intermediate PAC for criterion 31<sup>16</sup> and placement in the intermediate column was such that tables differed by more than one column in 12 cases (criteria 3, 7, 8, 9, 10, 11, 16, 18, 22, 24, 35 and 36).<sup>17</sup>

In general, the level of agreement between tables was higher for importance than for PAC. Consistency and overall evaluations are returned to in Section 3.2.2; the remainder of this section focuses on comments and discussion from the workshop.

The following points are based on the flipchart notes from table discussions and notes of the plenary discussion.

- Many criteria seem useful and important to DECC generally, but not directly to smart meters.
- Once the mass roll-out is under way, making valid comparisons against a control may become difficult or impossible.
- Should DECC be focused on the delivery of outcomes, not the paths/actions that suppliers are taking? In this model, DECC would be concerned with finer detail only if the roll-out was in some sense failing.
- It is difficult to separate out impacts due to smart meters – should we not be thinking more holistically?
- Collecting data on access rates (percentage of customers who have had successful installations, by customer type and property type) will be important as an early warning of any problems.
- The primary interest in many of the criteria lies with other policies (e.g. Smart Grids) and understanding this would be a separate dimension of work – energy retailers are not the only industry stakeholders, and potentially this workshop should be run with networks as well.
- Priorities change in terms of what to measure, how important it is, and how accurately to measure it once we are through the Foundation stage.
- For the Foundation stage, the priority should be getting the basic metering systems working, with progression into measuring changes in consumption as the programme moves into mass roll-out. But, if we need to measure and evaluate something later, we need to do it in the Foundation stage to establish measurement processes, learn about impacts and build up data on the historic baseline position.
- Prioritisation needs to take account of the trade-offs between the benefits of carrying out evaluation and the costs of measuring things. This may require a separate exercise to prioritise levels of accuracy, given that accuracy costs money.
- Different stakeholders would want to know different things or have different priorities for evaluation (e.g. the success of individual suppliers versus the industry as a whole), and the suppliers were less interested in the collation and distribution of data by the Government in relation to their market strategy.
- Individual suppliers may want to have measures of criteria where individual competition between suppliers is important, but not want to share this information (as there will be both successes and failures).

---

<sup>16</sup> 31 = Improvements in energy supply capacity in relation to specific products and services.

<sup>17</sup> 3 = The quality of information, feedback and support available to help consumers choose the most beneficial products and services in the roll-out as a whole. 7 = Improvements in customer convenience in the roll-out as a whole. 8 = Improvements in consumers' energy awareness/knowledge in the roll-out as a whole. 9 = Changes in consumers' energy-related behaviour in the roll-out as a whole. 10 = The uptake of smart-enabled products and services (other than new tariffs) in the roll-out as a whole. 11 = The uptake of new tariffs in the roll-out as a whole. 16 = Improvements in energy network capacity in the roll-out as a whole. 18 = Improvements in energy generation capacity in the roll-out as a whole. 22 = Improvements in industry cost-effectiveness in the roll-out as a whole. 24 = Improvements in customer satisfaction in relation to specific products and services. 35 = The uptake of smart-enabled products and services (other than new tariffs) in relation to specific consumer engagement approaches. 36 = The uptake of new tariffs in relation to specific consumer engagement approaches.

Capabilities on project:  
Building Engineering

- There could be customer resistance to collecting more data about them than suppliers do currently.
- Is the smart meter evaluation the right place to measure against *all* the criteria?
- In a number of cases (e.g. quality of supplier offerings), the suppliers would already be measuring against criteria for their own needs and they felt that the smart metering programme does not need to set up separate reporting criteria.
- For a number of criteria (e.g. quality of supplier offerings), one group felt it was not so important to the smart meter programme to deliver accuracy.

The following points are derived from comments made on individual criteria.

- Although the whole exercise was aimed at the Foundation stage, there were several comments that a criterion was more relevant to the long-term evaluation. This applied specifically to criteria 5, 14, 20, 22, 23, 29, 30, 33, 35 and 36. One concern was that evidence taken at Foundation stage, as the processes are being developed, could give a misleading impression of what would happen in the full roll-out when smart meters represent “business as usual”. In other cases, it was expected that benefits would appear only in the long term. More generally, suppliers saw the emphasis of the Foundation stage as being on getting the basic processes right. Also, there would be too few smart meters in the Foundation stage to evaluate broader outcomes such as energy supply efficiency.
- Some criteria were seen as difficult to measure – while this is not the first consideration, it will be important in final decisions over criteria. It was applied specifically to criteria 4, 9 and 20. One table commented that consumers’ energy awareness/knowledge was easier to measure as an aggregate effect (criterion 8) than in relation to specific consumer engagement approaches (criterion 34). One table commented that criterion 27 was easy to measure (as time of use of energy).
- Criteria 10, 14 and 22 were noted as being more important for suppliers than for DECC.
- One table felt that reference to “improvement” should be replaced by “changes” in criteria 12, 13, 14, 15, 16, 17, 18, 22 and 33. This is perhaps just a matter of how criteria are defined: the measure of performance would be change but the criterion for success would be improvement.
- There was some confusion over the distinction between supply, network and generation efficiency (criteria 12, 15, 17 and 30) in relation to the roll-out (or whether changes might be reliably attributed to smart meters). In any case, all four were placed similarly by each table except that one table thought 12 and 30 were both important and needed high quality measurement whereas the opposite was true of 14 and 17. One table commented that supply efficiency is important generally but not directly relevant to smart meters.
- Similarly there was some confusion over the distinction between supply, network and generation capacity (criteria 13, 16, 18 and 31) in relation to the roll-out. In any case, all three were placed similarly by each table except that one table thought 13 and 18 were important (16 and 31 were not) while 16 and 18 needed high quality measurement whereas 13 and 31 needed only intermediate quality measurement. Another table commented that capacity was not relevant because suppliers would supply what is required (this must be on the assumption that sufficient capacity will always exist). Another table pointed out the need to learn from the Foundation stage, even if capacity was not an explicit evaluation criterion.
- Criterion 2: suppliers had concerns about how information on different suppliers’ offerings might be used. They did not want a league table or information on competitive strategy to be shared (or used to create benchmarks). This is also true of other issues relating to marketing strategy.
- Criterion 8: the issue of persistence of impact was raised, and this may also apply to other criteria.
- Criteria 14 and 32: customer management efficiency is more important for the roll-out generally than for specific products and services.
- Criterion 15: the Low Carbon Network Fund was mentioned, presumably as a source of information.

Capabilities on project:  
Building Engineering

- Criterion 17: it was pointed out that criteria should relate to carbon emissions, not just energy efficiency, especially since “carbon is the key IA figure”. The DECC IA was also mentioned in relation to criterion 18 but the implication of the comment is less clear in this case.
- Criterion 25: one table thought the scope of energy-related behaviour needed to be clarified – whether it relates only energy use in the home or also for travel, etc.
- Criterion 28: one table commented that the quality of consumer trials is important for smart meters generally but not for specific products. Another table stated that this criterion is “High for roll-out critical issues in the Foundation stage”.
- Criterion 33: one table reported big differences in opinion, some thinking this criterion is not important in relation to smart meters but important for the industry and consumers.
- Criterion 34: the value of evaluating consumers’ energy awareness/knowledge was questioned on the grounds that it is not beneficial on its own – it matters only if it leads to changed behaviour.

### 3.2.2 Ratings at individual participant level

The exercise of table groups placing criteria on the grid was important in terms of generating discussion and understanding the thinking behind participants’ ratings of the criteria. For detailed analysis, we can return to the individual votes. The detailed findings are shown in tables and figures in Annex C.

Tables C1 and C2 show means and standard deviations of ratings of each criterion, for all participants and broken down by type of participant (suppliers vs others). The criteria are listed in rank order by mean rating of all participants, the top half of the ranks in the first table and the second half in the second table. Tables C3 and C4 show the equivalent data for ratings of precision/accuracy/confidence (PAC). In each case, the division of data into two tables is for convenience of presentation and does not imply a definitive difference between the two halves. The rank orders provide a starting point for deciding on which criteria to use and the quality of measurement that should be planned.

The standard deviations (SDs) in Tables C1-C4 potentially indicate the relative level of (dis)agreement but variation in SD is almost entirely (over 90%) accounted for by a non-linear (quartic) correlation between mean and SD: SDs are lower when the mean is very high or very low (this is a consequence of the short range of data points).

Figure C2 shows a plot of mean PAC rating against mean importance rating. Two main observations can be made on the figure.

- There is a non-linear (cubic) correlation between the two ratings. This implies that, in general, participants thought that if something was worth measuring, it was worth measuring well. This is more clear-cut at the extremes than in the middle of the range.
- Ratings relating to specific products and services (highlighted in orange) and specific consumer engagement strategies (highlighted in blue) are clustered in the middle of the importance range but cover the same range of PAC as ratings relating to the roll-out as a whole.

Suppliers and other participants differed in their ratings, more so for PAC (suppliers rated 0.19 higher on average) than for importance (suppliers rated 0.05 higher on average). In both cases, the difference increased with rating – see Figure C3. Criteria may be placed in rank order by differences in ratings – see Table C5. This table provides some indication of where DECC might need to take the initiative in relation to a particular criterion.

The above analysis of ratings is summarised by criterion in Tables 3.1-3.3, for criteria relating to the roll-out as a whole, criteria relating to specific products and services and criteria relating to specific consumer engagement approaches respectively.

Capabilities on project:  
Building Engineering

In each of these tables:

- quality refers to the precision, accuracy or confidence of measurement;
- “Table result” shows, for Tables A-C, whether the criterion was placed high (H), medium (M) or low (L);
- ranks are based on all 36 criteria and all participants, and a high rank by mean indicates a high rating (i.e. a criterion ranked 1 for importance is the most important criterion);
- where more than one criterion had the same rating, equal ranks as denoted by the = sign;
- rank by diff. (difference) refers to the rank order of criteria by difference between ratings by suppliers and others (a high rank means suppliers gave higher ratings).

**Table 3.1** Main findings on evaluation criteria relating to the roll-out as a whole

Criterion	Importance ratings			Quality ratings		
	Table result	Rank by:		Table result	Rank by:	
		mean	diff.		mean	diff.
1. The quality of <i>smart metering installation, services and support</i>	HHH	1=	19=	MHH	12	3=
2. The quality of <i>energy supply offerings</i>	HMH	8=	12=	MMH	20=	21=
3. The quality of <i>information, feedback and support available to help consumers choose the most beneficial products and services</i>	HHH	5=	12=	MLH	19	8=
4. The quality of <i>data security and privacy</i>	HHH	1=	19=	HMH	6=	3=
5. The degree of <i>competition in relevant markets</i>	MHH	18=	19=	MLL	34	28=
6. Savings on <i>domestic energy bills</i>	HHH	14=	31=	HHH	6=	3=
7. Improvements in <i>customer convenience</i>	HHH	8=	1	MLL	31=	3=
8. Improvements in <i>consumers' energy awareness/knowledge</i>	HMM	18=	6=	MLH	28=	15=
9. Changes in <i>consumers' energy-related behaviour</i>	HHH	3=	23=	HLH	8=	32=
10. The <i>uptake of smart-enabled products and services</i> (other than new tariffs)	HHH	28	12=	MML	23=	21=
11. The <i>uptake of new tariffs</i>	MHH	8=	12=	MHL	23=	11=
12. Improvements in <i>energy supply efficiency</i>	MLH	22=	23=	MMH	22	15=
13. Improvements in <i>energy supply capacity</i>	MHL	22=	31=	MML	31=	21=
14. Improvements in <i>customer management efficiency</i> (e.g. process simplification)	HLH	14=	1	HMH	16=	1
15. Improvements in <i>energy network efficiency</i>	MLL	35	36	MLL	35	21=
16. Improvements in <i>energy network capacity</i>	MLL	33=	34	MHL	16=	35
17. Improvements in <i>energy generation efficiency</i>	MLL	36	6=	MML	28=	28=
18. Improvements in <i>energy generation capacity</i>	MHL	33=	12=	MHL	27	15=
19. The <i>quality of consumer trials</i>	HHH	5=	28=	HHH	3=	32=
20. Reduced <i>overall energy demand</i>	HHH	3=	23=	HHH	1	21=
21. <i>Smoother time profile</i> of domestic energy use	HMM	29=	23=	HHH	16=	21=
22. Improvements in <i>industry cost-effectiveness</i>	MLL	32	12=	MHL	20=	3=

Capabilities on project:  
Building Engineering

**Table 3.2** Main findings on evaluation criteria relating to specific products and services

Criterion	Importance			Quality		
	Table result	Rank by:		Table result	Rank by:	
		mean	diff.		mean	diff.
23. Savings on <i>domestic energy bills</i>	HHH	8=	28=	HHH	2	15=
24. Improvements in <i>customer satisfaction</i>	HHH	14=	3	MLL	28=	8=
25. Changes in <i>consumers' energy-related behaviour</i>	HMH	8=	28=	MMH	8=	28=
26. Reduced <i>overall energy demand</i>	HHH	7	19=	HMH	3=	36
27. <i>Smoother time profile</i> of domestic energy use	HHM	24=	35	HHM	13=	32=
28. The <i>quality of consumer trials</i>	HHH	14=	23=	HMH	3=	28=
29. The degree of <i>competition in relevant markets</i>	MHH	20	6=	LLL	36	13=
30. Improvements in <i>energy supply efficiency</i>	LLH	29=	2	MMH	13=	19
31. Improvements in <i>energy supply capacity</i>	LLM	29=	33	MMM	11	13=
32. Improvements in <i>customer management efficiency</i> (e.g. process simplification)	LLL	27	10	MLL	23=	2
33. Improvements in <i>industry cost-effectiveness</i>	LMH	21	12=	MMH	8=	8=

**Table 3.3** Main findings on evaluation criteria relating to specific consumer engagement approaches

Criterion	Importance			Quality		
	Table result	Rank by:		Table result	Rank by:	
		mean	diff.		mean	diff.
34. Improvements in <i>consumers' energy awareness/knowledge</i>	HMH	13	4	HMH	15	11=
35. The <i>uptake of smart-enabled products and services</i> (other than new tariffs)	MMH	24=	5	LHL	23=	20
36. The <i>uptake of new tariffs</i>	MHL	26	6=	MHL	16=	21=

The criteria given a top 8 (or 8=) importance rank, or placed in a high importance box by all three tables, may be seen as the criteria identified by the workshop as the most important. These criteria all relate to the roll-out as a whole, except in five cases where a criterion is important for the roll-out as a whole and the equivalent criterion is important for specific products and services. There were no cases of a criterion being important for specific products and services but not important for the roll-out as a whole, and no cases of a criterion being important for specific consumer engagement approaches.

The following important criteria relate directly to consumer benefits (and therefore, indirectly, benefits for suppliers).

6. Savings on domestic energy bills in the roll-out as a whole. / 23. Savings on domestic energy bills in relation to specific products and services.

20. Reduced overall energy demand in the roll-out as a whole. / 26. Reduced overall energy demand in relation to specific products and services.

7. Improvements in customer convenience in the roll-out as a whole. / 24. Improvements in customer satisfaction in relation to specific products and services.

Capabilities on project:  
Building Engineering

Two important criteria relate to consumers' behavioural response, by which they could realise benefits to themselves.

9. Changes in consumers' energy-related behaviour in the roll-out as a whole. / 25. Changes in consumers' energy-related behaviour in relation to specific products and services.

The following criteria relate to the services provided to consumers and can be considered a benefit to them but also to the supplier that wins or retains business because of the service.

1. The quality of smart metering installation, services and support in the roll-out as a whole.
2. The quality of energy supply offerings in the roll-out as a whole.
3. The quality of information, feedback and support available to help consumers choose the most beneficial products and services in the roll-out as a whole.
10. The uptake of smart-enabled products and services (other than new tariffs) in the roll-out as a whole.<sup>18</sup>
11. The uptake of new tariffs in the roll-out as a whole.

One criterion might also be considered in this same category but it is a service that is generally invisible to the consumer.

4. The quality of data security and privacy in the roll-out as a whole.

Finally, two important criteria relate to how trials are conducted.

19. The quality of consumer trials in the roll-out as a whole. / 28. The quality of consumer trials in relation to specific products and services.

In similar fashion, the least important criteria may be identified. In this case, most relate to industry efficiency and capacity.

15. Improvements in energy network efficiency in the roll-out as a whole.
16. Improvements in energy network capacity in the roll-out as a whole.
17. Improvements in energy generation efficiency in the roll-out as a whole.
18. Improvements in energy generation capacity in the roll-out as a whole.
30. Improvements in energy supply efficiency in relation to specific products and services.
31. Improvements in energy supply capacity in relation to specific products and services.
22. Improvements in industry cost-effectiveness in the roll-out as a whole.
32. Improvements in customer management efficiency (e.g. process simplification) in relation to specific products and services.

In addition, one relates to load shifting. This was rated slightly higher in relation to specific products and services, perhaps indicating that it was seen as relating to tariffs rather than smart meters as such.

21. Smoother time profile of domestic energy use in the roll-out as a whole.

---

<sup>18</sup> Criterion 10 is unusual in being placed in a high importance box by all three tables and yet having a low overall importance rank. The table decisions appear to have been biased towards those who thought it was important.

Capabilities on project:  
Building Engineering

## 4 Roll-out activities

The presentation of possible roll-out activities<sup>19</sup> drew relatively little discussion but the following points were made.

- The baseline for the evaluation needs to be defined so that everyone knows what any changes are being measured against.
- The Annual Quantity database (of consumption) is a macro-dataset – useful for looking across regions but may not be very helpful for understanding impacts of individual interventions. Also, it will be important to extract these data at the right point in processing, e.g. before/after temperature correction.
- ELEXON's profile sampling meter base may be relevant, providing accurate, representative data. The timing of when a smart meter is installed will affect the apparent impact, so data on the timing of installations is required.
- A supplier noted that, as new products come on the market, there will be a need to compare within smart meter products and services.
- There will be other effects on baseline energy demand, underlying trend use changes, and effects of energy prices (for example) which will all affect the dynamics of the baseline. It is important to have appropriate control groups.
- Two comparisons were distinguished: between smart and non-smart (on installation); and when introducing a further intervention such as a time of use tariff.
- There was a discussion about whether the non-smart meter baseline was usable in terms of quality; it was suggested that if the sample size is big enough and followed over a long enough period, this could deal with the noise in non-smart meter data.

---

<sup>19</sup> Slides 21-22 in Supplementary Document C.

## 5 Evaluation good practice

### 5.1 Method

This part of the workshop<sup>20</sup> began with a presentation of some basics of good practice in evaluation – a quick overview to remind participants of the range of issues. This used headings of design, sampling, recruitment, data collection, customer feedback and data analysis. The main points from the slides were reproduced on A3 sheets, attached to the walls – one sheet per heading (see Supplementary Document E).

Participants then individually contributed suggestions as to the possible practical problems with following each aspect of good practice. This was done by writing the problem on sticky note and placing it on the appropriate poster. Once everyone had finished writing, there was a plenary discussion of one poster at a time; the potential problems were summarised and – where possible – solutions suggested.

In the report of findings from these sessions, information has been used from the sticky notes, flipchart notes made during the discussion and more extensive notes from the meeting note-taker. Some items have been moved between headings in order to rationalise the report. In other cases, the discussion may not be placed strictly under the correct heading but has not been moved, so as to maintain the integrity of the flow of issues.

### 5.2 Findings

#### 5.2.1 Initial discussion

Following the presentation, the following points came out of the discussion.

- Getting the Foundation right, as the platform for products and services, is fundamental. Development and assessment of products and services then follows on top of that.
- Although the immediate supplier priority is gaining access to customers, and getting meters working and interoperable with other equipment, in public policy terms the focus has to be on whether benefits are being realised. The Programme has to know about this.
- A supplier said he wanted to understand what would give DECC confidence that benefits were being delivered. However, there was commercial sensitivity over publicising the detailed results of trials.
- More generally, there is a need to recognise commercial sensitivities.

#### 5.2.2 Evaluation design

Validity is a key issue – internal validity (are we measuring the right things, are we measuring energy use accurately enough and can we attribute any change to a smart meter) and external validity (are the results representative of the general population)? Internal validity is compromised because the main “treatment” (smart meter installation) cannot be restricted, if customers request it, and neither can some other important variables (e.g. customer choice to change supplier). External validity is also compromised by self-selection (at least in the Foundation stage), i.e. opting in to having a smart meter.

One approach to the problem of self-selection would be to base the evaluation of standard installations on a selection of customers chosen on an opt-out basis (opt-in may be required for at least some additional smart-enabled products and services). This approach would require a control group and there is no basis for preventing the control group having a smart meter installed, as noted above, so the control group would need to be made large

---

<sup>20</sup> Slides 23-38 in Supplementary Document C. Participants had a paper copy of the presentation for easy reference during the exercise and they were reminded of this at the start of the presentation.

Capabilities on project:  
Building Engineering

enough to allow for these losses. It also needs to be understood that an opt-in sample gives valid findings for people who are choosing the intervention – the important point is not to generalise those findings to people who have not chosen the intervention without first accounting for the likely biases. The timing of installations may be a critical factor in relation to impact, i.e. whether the installation is carried out under a universal mandate or by supplier/consumer choice. Opt-in studies can also help to identify interested customer segments and understand what gets them to “sign up”.

The fact that customers have choices extends beyond the basic installation. Householders may undertake various installations or change their behaviour in ways that will create ‘noise’ in the data if they are not measured or observed. To some extent, the order in which changes occur will help in the interpretation of data – sometimes there may be a causal relationship (e.g. a smart meter could enhance the customer’s understanding of demand and lead to PV installation) but this would not apply in all cases. Energy suppliers lack wider data on household characteristics and other variables (e.g. installation of PV or a new boiler around the same time as the smart meter). This may be because the information is inaccurate or split between two suppliers (where gas and electricity are supplied by different companies), or held by another party with data security restrictions, or simply because nobody records the information. If such data were required, suppliers would need to undertake dedicated data collection as part of the roll-out generally or specific trials.

The implied need for a wide range of measurements and observations, to remove noise and clarify the impact of interventions, suggests that only small samples will be feasible with the full range of variables. Close matching of samples would mean that much smaller sample sizes would be needed in order to achieve statistically significant effects. Design decisions over matching will therefore have critical implications for sampling, recruitment and measurement.

It was also suggested that installing a smart meter and not telling the household about it (e.g. by using the business-as-usual explanation the meter was “due to be changed anyway”) would become a non-compliant installation, given the introduction of the installation Code of Practice (i.e. meeting the needs of consumer protection or compliance with other Programme requirements). This would place a restriction on the definition of control groups. DECC therefore needs to clarify the legitimacy of a smart meter installation not being communicated to the householder.

Commercial sensitivity was another recurring theme. During the Foundation stage, there is no obligation on suppliers to do anything in relation to the roll-out. There was some doubt that suppliers would share information on early mover installations in these circumstances: it is the suppliers’ choice to invest or not and they would naturally wish to have restrictions on what ‘intellectual capital’ is shared with those who choose not to invest. A particular issue is that customer segmentation varies between suppliers and this segmentation is itself commercially sensitive, so there may be barriers to breaking down data by customer segment.

However, in practice, commercial sensitivity will depend on the type of information, the format and time frames in which it is provided and to whom it is provided. In general, suppliers were (unsurprisingly) more comfortable sharing information with Government than with competitors. Suppliers can report feedback on commercially sensitive interventions in confidence to DECC/Ofgem (at a regional/national level). DECC/Ofgem can evaluate whether particular activities deliver benefits. This is likely to apply to some extent throughout the roll-out, not just in the Foundation stage.

The answer to the problem of commercial sensitivity therefore lies in prior agreement on the basis on which data can be shared. In particular, DECC needs to be clear about what it wants, and develop a plan for the information required and who will be able to have access to it. The need for DECC to specify data requirements was also emphasised in other ways.

- Suppliers may want data for one purpose and DECC for another – suppliers need prior agreement if data are to be provided to DECC.

Capabilities on project:  
Building Engineering

- In relation to Foundation trials, some suppliers were concerned that it may already be too late – trials take two years to plan and execute.
- If already started, Foundation trials may be difficult to change to collect data deemed relevant by DECC.
- A specification needs to cover what should be collected as standard/routinely (and how often) and what should be collected in 'deep dives'.
- There needs to be clarity over how Foundation requirements translate into enduring obligations.

A supplier made the point that confidence in the evaluation may be enhanced by linking it to benefits identified in the DECC impact assessment, and that doing this would tend to focus the evaluation on energy consumption and load shifting.

The over-arching aim was seen as a need to demonstrate an *achievable* level of savings, during Foundation. This requires good baseline data and there is limited evidence on this, although it does exist (electricity profiling). Suppliers may simply have to confirm what baseline they have used and state any caveats with the information.

### 5.2.3 Sampling and recruitment

Suppliers felt that DECC only needs to evaluate impacts across the total population (in line with the benefits case), but suppliers may want to target narrower segments and particular cause-effect relationships. They need to understand what does or does not work for specific segments and to understand the mechanisms by which particular interventions have particular effects. This is related essentially to competitive advantage but it may also bias overall results – this would be in addition to any bias due to opting in. Also, MPs may want to know what is happening in their constituencies, creating a need for geographic breakdown of samples. However, there is a balance between cost and benefit in terms of the level of detail, and not over-complicating the Foundation stage.

One issue around population segments is that, in the Foundation stage, there is likely to be a tendency for suppliers to avoid hard-to-install properties, prepayment customers and consumers with less benefit/engagement (e.g. those with low energy use). Again, this would distort results in relation to expectations for the full roll-out; we need to understand whether the early mover installations are representative. Two approaches were identified for dealing with this:

- accept that lessons need to be learned from easier installations before moving on to the more difficult ones;
- expect some difficult installations to be made in the early Foundation phase, to ensure that the practical issues are fully understood and resolved (e.g. communications solutions for flats where the meter is some distance from the dwelling) and differences in impact are identified.

The latter could target all homes in a limited geographical area and may require Government funding support. The strategy needs to be decided as soon as possible.

Other points made were:

- standards for statistical robustness (e.g. level of confidence) need to be agreed with DECC in advance;
- representative sampling will simplify statistical analysis but segmentation choice will vary between suppliers;
- matching data across trials/samples will require timing data (e.g. in relation to effects of weather and cultural events such as August holidays and Christmas);
- meaningful data will be at half-hourly level and suppliers understand that they will need customer permission to use such data in evaluation;

Capabilities on project:  
Building Engineering

- an upfront and honest approach to recruitment sampling will generate a better response from consumers – to do otherwise runs the risk of the roll-out being halted because of privacy issues and campaigns.

#### 5.2.4 Data collection

There was a general acceptance that consistency of data collection methods and standards is important and that these therefore need to be centrally agreed. While methods that are already established may appear to offer least cost, conversion from methods actually being used by suppliers may prove difficult and expensive. This leads back to the common theme that DECC, in consultation with all those interested, should provide clarity on what data and/or trials it needs, for what purposes. There was particular concern that any data required centrally need to be “properly designed and clearly understood”. A key aspect of achieving consistency was perceived to be whether there will be a central data collection point.

The need for data specification was partly related to questions over the level at which DECC needs to evaluate. If the roll-out as a whole (or the roll-out by a particular supplier) is delivering a positive impact overall (e.g. 5% savings in energy demand), does DECC need to know any more? There may also be practical issues here, in terms of what level of detail is actually achievable during the Foundation stage. Nevertheless, in order to be confident of the overall impact, a wide range of data may be needed (e.g. baseline data and data to control for general trend changes and disruptive events).

A second major theme was a need to be selective – it will not be feasible to measure everything and there is a trade-off between doing a few things well or many poorly (quality vs quantity). In addition, suppliers may already have determined clear trial objectives and data requirements, and it could be difficult to add to these. However, it will be important to have the data required to pick up signals from the background noise, and distinguish minor and major effects.

#### 5.2.5 Customer feedback

This is a specific aspect of data collection and the general points from the previous section therefore also apply.

The issue of consistency of methods was discussed in relation to the potential value of a short harmonised questionnaire to use in all data collection exercises (possibly by self-completion). Questions would need to be asked always in the same sequence, before other, non-standard questions. The standard questions would then influence responses to the non-standard questions (which might be the main focus of the suppliers’ interest).

It also matters who administers a questionnaire, so there would need to be a standard type of interviewer (supplier, Government, University, etc.). This was seen as important in relation to the potential for bias in customer responses (e.g. over-estimating / over-reporting changes in behaviour) or willingness to provide personal household data at all. A market research organisation, independent of the industry and Government, may get truer perspective than supplier-sponsored research. It was also felt that consumer groups can (and probably will) undertake these exercises anyway, the implication perhaps being that DECC and suppliers do not need to worry about it. This raises issues over linking datasets so that either the energy supplier has feedback data or the other organisation has consumption data, at household level. Data privacy was also seen more generally as a barrier, particularly if customers have the right to opt out of data linking (and there being a risk of this being the norm).

Other specific points made were:

- the importance of common central messages to mitigate programme risks that could distort impact and take-up (e.g. concerns over EMF, perception that the programme is too expensive);
- feedback is more important from consumers where it has *not* worked for them;

Capabilities on project:  
Building Engineering

- there may be customer fatigue with surveys and the number of interventions;
- there needs to be a plan for how (and how often) to capture relapse in behaviour – trailing off of understanding, engagement and behaviour change.

### 5.2.6 Data analysis

A key concern here was the mere availability of data (e.g. because of data protection) and the massive storage, processing and retrieval challenge. Four specific issues were raised.

- There were concerns over “certainty of data” during the early period (with an expectation that pre-DCC processes would be “more cumbersome” and have more “teething problems”) leading to customer service issues.
- Suppliers are building complex IS Systems and it will be challenging to get data out of these, e.g. correlating product sales with consumption change impacts, if requirements are not specified early enough. However, the DCC may make things easier, by providing accurate information on consumption (likely to be from 2014).
- Granularity of control group data may be a problem (e.g. lack of manual meter reads).
- If suppliers rely on customers opting in to have access to their detailed data, this would make it impossible to analyse properly, and would also introduce a bias between different customer segments as levels of opting in would vary between customer types.<sup>21</sup>

An important general point concerned who should carry out the analysis. This related to the conduct of the analysis but, more importantly, to independence and trust in the findings. Data collection/evaluation by an independent body or Government may be better received by consumers. Similarly, it is important to be clear about the audience for the analysis (e.g. suppliers, DECC, Ofgem, the public) and what data/analysis can be aggregated for public consumption or used internally (by suppliers or Government) for refinement of the Programme.

A particular challenge for the analysis will be multiple policies impacting on the same outcomes. DECC forecasts energy usage and anticipated energy savings associated with each policy. This is a starting point for assessing if total consumption changes in line with DECC’s expectations but the analysis needs to take account of the timing of other interventions (e.g. Green Deal installations). It will also need to take into account varying approaches to the Foundation stage (actions trialled in the Foundation stage could create bias later) and to monitor additional third party interventions.

A further source of variation that needs to be controlled for (if it is not to be a source of noise in the data) is consistency of message in customer contacts. Contacts may go to different points within a supplier organisation (e.g. general call centres, call centres dedicated to smart meters, or the installers), resulting in different messages. Suppliers saw it as impossible to standardise supplier contacts for the purposes of evaluation.

More generally, there is a moving baseline of customer knowledge and understanding: these are constantly changing and the analysis will need to control for this.

The discussion returned to the theme of needing a clear specification from DECC about reporting requirements and hence data quality: what is the scope and what are the “targets” and to what resolution should they be assessed?

Finally, a very specific concern was how to identify the cause for ‘hot-spots’ of increased consumption.

---

<sup>21</sup> This does not apply if the provision of the data to Government is a “regulated duty” and it is likely that DECC will want to regulate for data access in some form, at some point. DECC statisticians have just moved to using legislation to underpin collection of consumption data for NEED.

Capabilities on project:  
Building Engineering

## 6 Organisation of the evaluation

### 6.1 Introduction

A brief presentation<sup>22</sup> addressed the overall organisation of the evaluation – how standards will be set and assured to ensure a minimum level of quality without suppressing creativity and innovation. It was emphasised that the slides did not represent any policy commitment or plan in terms of influencing roll-out, but there is a need to find a sufficiently robust solution. Who will decide what is good enough? Is there a role for an expert panel to advise on suppliers' approaches, or should there be only ex-post expert peer review? The example model shown in the slides was not specifically discussed.

In practice, many of the points that might have been discussed here had already been made during earlier sessions.

### 6.2 Findings

Suppliers made the following points.

- Suppliers' strategies will be influenced by practical and commercial factors, e.g. whether they initially avoid difficult properties and customers, to help them hit targets; or conversely, how they will prioritise customers where debt management is an issue, as they will want to drive out avoidable debt as quickly as possible.
- Suppliers will take their own roll-out approach, making their own choices. But in the context of evaluation, there is a need to have a common framework for measuring benefits.
- Energy efficiency is dependent on many factors that are not the main focus of suppliers' interests (although varying across suppliers) and no energy-saving targets have been agreed.
- Suppliers' benefits also need to be taken into account.
- If the roll-out is not successful in terms of energy-saving and energy efficiency objectives are not achieved, then suppliers will need to undertake more activities to address this.
- If there is no information on benefits, NAO will scrap the programme.
- Trials would be about "excellent customer service", not exactly the same things that the Programme is interested in. It is vital that DECC designs its requirements carefully in view of this.
- We need to keep it simple, as it is already very late to be introducing evaluation plans for the Foundation stage. DECC needs to identify quickly what needs to be measured and then "get on with it", and do trials separately. Suppliers need to know soon what information is needed for evaluation.
- One supplier is already working with partners (National Housing Federation, Cabinet Office Behavioural Insight Team, etc.) and will share those trials with DECC. On "other stuff", we should "let suppliers deal with it".

Other participants made the following points.

- DECC would want to know if the approaches were comparable, and having similar impacts.
- It will be impossible to come up with agreed standards across different suppliers' roll-out packages.
- We should be collecting data on aggregate impacts, and pursuing Government/academic partnerships (with supplier inputs) to develop scientific trials, rather than trying to bend suppliers' roll-out activities to provide comparable situations for evaluation, treating them as experiments.
- Where specific interventions are the focus, this should be treated as more academic research, but with supplier collaboration.

---

<sup>22</sup> Slides 39-42 in Supplementary Document C.

Capabilities on project:  
Building Engineering

## 7 Remaining issues

### 7.1 Introduction

The final session was an opportunity for suppliers to say which aspects of evaluation they were likely to find most difficult and what support they are most likely to need (e.g. what expertise, input, topics or types of person/organisation). The other parties at their table were available to help them to refine, detail and clarify the need for support. This was to assist DECC in understanding suppliers' issues, to prepare the ground for setting out reasonable expectations. The difficulties were written on sticky notes and passed to AECOM for consideration at report stage.

### 7.2 Findings

The following can be seen as difficulties in the sense intended for this session.

- Customer/market confidence: DECC/industry uncertainty could drive loss of credibility.
- There may be too much demand for smart meters, leading to customers being disappointed. A solution to this was noted as "Publish a realistic picture of what savings customers can achieve".<sup>23</sup>
- There may be restrictions on suppliers' ability to talk to customers about energy-saving products at the point of smart meter installation and this will impact consumption reduction.
- Suppliers may have limited access to smart meter data for the purpose of evaluation. A solution to this was noted as "DECC should not restrict suppliers' ability to collect and use data".

Other comments made at this stage were as follows.

- Focus on a simple business-as-usual solution at least cost, especially during the Foundation stage.
- Agree time-based 'targets'.
- Do not constrain a supplier-led roll-out.
- Evaluation must ensure that the Programme is capable of showing that the smart meter roll-out is working. It must be able to highlight problems and ensure that customer confidence remains.
- Avoid duplication of effort.

---

<sup>23</sup> This appears to relate to a different problem, i.e. that customers may have an unrealistic view of what savings can be achieved.

Capabilities on project:  
Building Engineering

## 8 Conclusions

### 8.1 Introduction

The preceding chapters set out the findings by workshop session. This chapter draws those findings together to form overall conclusions by theme. These should be seen as exactly what they are – conclusions from a workshop with a particular group of participants. While not definitive or comprehensive, they are provided to assist DECC in understanding suppliers' issues and more broadly to inform decisions on how to take forward the roll-out evaluation.

### 8.2 The purpose of the evaluation

Evaluation of the smart meter roll-out is important:

- to give greater certainty over the best approach and the likely impacts (especially on householder behaviour);
- to provide feedback during the roll-out, to improve the roll-out process and provide early warning of risks;
- for suppliers to maximise the benefit to themselves, customers and UK commitments to reduce energy demand;
- to give the public (and NAO) confidence that the roll-out is of value and being conducted fairly and efficiently by Government and the supply industry (thus maintaining credibility of the Programme);
- to benefit from overlaps with other policies and programmes;
- because there is a commitment to do it.

### 8.3 Evaluation criteria

There are many possible evaluation criteria (i.e. ways of judging the success of the roll-out). They can be characterised by the potential beneficiaries (customers, suppliers, etc.), the scope of the evaluation, the roll-out process and impact, the source of data, and synergy with other actions. The findings, summarised in Tables 3.1-3.3, indicate how suppliers and others rate the *importance* of various criteria and the *quality* of measurement (precision, accuracy or confidence) that would be appropriate for each. Five general points can be made from the ratings data.

- The rank order of ratings provides a starting point for deciding which criteria to use and the quality of measurement that should be planned.
- In general, participants thought that if something was worth measuring, it was worth measuring well.
- Ratings of importance were highest for the following criteria relating to the Foundation stage roll-out as a whole:
 

<ul style="list-style-type: none"> <li>➢ savings on domestic energy bills;</li> <li>➢ reduced overall demand;</li> <li>➢ customer convenience/satisfaction;</li> <li>➢ changes in consumer behaviour;</li> <li>➢ quality of consumer trials;</li> <li>➢ quality of smart metering installation, services and support;</li> </ul>	<ul style="list-style-type: none"> <li>➢ quality of energy supply offerings;</li> <li>➢ quality of information, feedback and support available to help consumers choose the most beneficial products and services;</li> <li>➢ uptake of new tariffs and other smart-enabled products and services;</li> <li>➢ data security and privacy.</li> </ul>
--	---
- The first five of these also attracted the highest importance ratings in relation to specific products and services. In general, criteria relating to specific products, services and consumer engagement strategies had intermediate importance ratings.
- Suppliers tended to give higher ratings than other participants for quality requirement and slightly higher for importance. In both cases, the difference increased with rating. Criteria may be placed in rank order by these differences, indicating where DECC might need to take the initiative in relation to a particular criterion.

Capabilities on project:  
Building Engineering

The main points from the discussion and from comments made on individual criteria are as follows.

- Priorities change after the Foundation stage: criteria 5, 14, 20, 22, 23, 29, 30, 33, 35 and 36 were specifically noted as being more relevant to long-term evaluation (e.g. because benefits would take time to appear, or there would be too few smart meters to evaluate outcomes, or early evidence could give a misleading impression of what would happen in the full roll-out). But, if we need to measure and evaluate something later, we need to do it in the Foundation stage to establish measurement processes, learn about impacts and build up data on the historic baseline position (accepting that impacts may differ between the Foundation and main roll-out).
- Criteria 4, 9, 20 and 34 were noted as difficult to measure – while this is not the first consideration, it will be important in final decisions.
- Persistence of impact needs to be considered.
- The Low Carbon Network Fund might provide evaluation data on network-level impacts.
- Criteria should relate to carbon emissions, not just energy use.
- Consumers' energy awareness/knowledge matters only if it leads to changed behaviour.
- Criteria should relate specifically to smart meters, as distinct from being generally useful and important.
- Suppliers felt that DECC only needs to evaluate impacts across the total population (in line with the benefits case) and the delivery of outcomes, not the paths/actions that suppliers are taking. In this model, DECC would be concerned with finer detail only if the roll-out was in some sense failing. Nevertheless, in order to be confident of the overall impact, a wide range of data may be needed, and specific areas of difficulty identified.
- Suppliers, in contrast, are focused on "excellent customer service" and competitive advantage; they may want to understand what does or does not work for specific segments and to understand the mechanisms by which particular interventions have particular effects. This may also bias overall results.
- Suppliers may want to have criteria where competition between suppliers is important (e.g. quality of offerings), but not want to share this information (as there will be both successes and failures) or information on their market strategy. In general, their focus is their own success rather than that of the industry as a whole.
- MPs may want to know what is happening in their constituencies, creating a need for geographic breakdowns.
- The percentage of customers who have had successful installations (by customer and property type) will provide early warning of problems.
- Other industry stakeholders should be consulted on criteria that affect them.
- Prioritisation needs to take account of the trade-offs between the benefits of carrying out evaluation and the costs of measuring things. This may require a separate exercise to prioritise levels of accuracy.

## 8.4 Organisation of the evaluation

Organisation of the evaluation is about how DECC, suppliers and other parties will work together to deliver the evaluation, avoiding duplication of effort and setting and assuring standards to ensure a minimum level of quality without suppressing creativity and innovation by suppliers.

For the suppliers, getting the Foundation right, as the platform for products and services, is fundamental. The immediate supplier priority is gaining access to customers, and getting the basic metering systems working and interoperable. Development and assessment of products then follows on top of that, with progression into measuring changes in consumption as the programme moves into mass roll-out. Nevertheless, in public policy terms, the focus has to be on whether benefits are being realised. In this context, confidence in the evaluation may be enhanced by linking it to benefits identified in the DECC impact assessment.

Capabilities on project:  
Building Engineering

Commercial sensitivity was a recurring theme (e.g. around publication of the detailed results of trials). During the Foundation stage, there is no obligation on suppliers to do anything in relation to the roll-out. There was some doubt that suppliers would share information on early mover installations in these circumstances: it is the suppliers' choice to invest or not and they would naturally wish to have restrictions on what 'intellectual capital' is shared with those who choose not to invest. A particular issue is that customer segmentation varies between suppliers and this segmentation is itself commercially sensitive, so there may be barriers to breaking down data by customer segment.

However, in practice, commercial sensitivity will depend on the type of information, the format and time frames in which it is provided and to whom it is provided. In general, suppliers were (unsurprisingly) more comfortable sharing information in confidence with Government (at regional/national level) than with competitors. DECC/Ofgem could then evaluate whether particular activities deliver benefits. This is likely to apply to some extent throughout the roll-out, not just in the Foundation stage.

Commercial sensitivity may therefore be addressed through agreement of a plan (drafted by DECC) for the data required, who will be granted access and on what basis. This is part of a broader message from the workshop, that suppliers will have their own priorities for the Foundation (and full roll-out), in relation both to executing the roll-out and gathering data. Suppliers are therefore looking to DECC (in consultation with all interested parties) to specify what additional activities (if any) and what data will be expected and what would give DECC confidence that benefits were being delivered and consistently assessed: a common framework for measuring benefits. At the same time, there was some doubt over the feasibility of agreeing common standards across different suppliers' roll-out packages. Nevertheless, participants thought that DECC's specification should include:

- what baseline and post-roll-out data should be collected as standard or routine (and how often) and what should be collected in more limited in-depth data collection;
- the reasons the data are needed;
- any design requirements or standard methods that have to be adhered to;
- standards for data quality and statistical robustness (e.g. level of confidence);
- reporting requirements, targets and the resolution with which should they be assessed;
- how Foundation requirements will translate into enduring obligations.

This was seen as urgent, to achieve consistency of methods and data and for suppliers to know what is expected of them. In relation to Foundation trials, some suppliers were concerned that it may already be too late – trials take two years to plan and execute. If already started, trials may be difficult to change. In this context, suppliers expressed a need to "keep it simple" with, as far as possible, a business-as-usual solution at least cost in the Foundation. However, there was also concern that, once the mass roll-out is under way, making valid comparisons against a control may become difficult or impossible.

Targets should also be translated into information for customers, to retain their confidence and avoid unnecessary disappointment (e.g. if suppliers cannot meet demand for smart meters, or if customers have an unrealistic expectation of what savings they can achieve).

Other points made were as follows.

- A key aspect of achieving consistency is whether there will be a central data collection point, such as the DCC.
- It is important to have common and consistent central messages, to mitigate programme risks that could distort impact and take-up (e.g. concerns over EMF or perception that the programme is too expensive).
- The smart meter evaluation may not be the right mechanism to measure against *all* the criteria.

Capabilities on project:  
Building Engineering

- DECC should collect data on aggregate impacts and pursue partnerships (e.g. with academic researchers, and with supplier input) to develop scientific trials, rather than trying to bend suppliers' roll-out activities to provide comparable situations for evaluation, treating them as experiments.

## 8.5 Execution of the evaluation

### 8.5.1 Evaluation design

External validity is compromised by self-selection, i.e. opting in to having a smart meter. This would largely be overcome by basing the evaluation of standard installations on a selection of customers chosen on an opt-out basis (opt-in may be required for at least some additional smart-enabled products and services). This approach would require a control group and there is no basis for preventing the control group having a smart meter installed, so it would need to be large enough to allow for these losses. The possibility was discussed of creating control groups by installing a smart meter and not telling the household (e.g. by presenting the installation as business-as-usual meter replacement). There was concern that this would be a non-compliant installation under the installation Code of Practice; DECC therefore needs to clarify the legitimacy of this approach.

An opt-in sample does give valid findings for people who are choosing the intervention – the important point is not to generalise those findings to people who have not chosen the intervention without first accounting for the likely biases. Opt-in studies can also identify interested customer segments and find out what gets them to 'sign up'.

Internal validity is compromised because smart meter installation cannot be restricted and neither can other consumer choices (e.g. to change supplier, undertake installations or change behaviour) that will create noise in the data if they are not measured or observed. To some extent, the order in which changes occur will help in the interpretation of data – sometimes there may be a causal relationship (e.g. a smart meter could enhance the customer's understanding of energy demand and lead to installation of PV or a new boiler) but this would not apply in all cases. Energy suppliers lack wider data on such variables, which would be needed to interpret such extraneous changes. This may be because the data are inaccurate, split between gas and electricity suppliers, held by another party with data security restrictions, or because nobody records the information. If such data were required, suppliers would need to undertake dedicated data collection (in the roll-out generally or in specific trials).

The implied need for a wide range of measurements and observations to remove noise and clarify the impact of interventions suggests that only small samples will be feasible with the full range of variables. Close matching of samples would mean that much smaller sample sizes would be needed to achieve statistically significant effects. Design decisions over matching will therefore have critical implications for sampling, recruitment and measurement.

### 8.5.2 Sampling and recruitment

Suppliers' strategies will be influenced by practical/commercial factors. In the Foundation stage, there is likely to be a tendency to avoid hard-to-install properties, prepayment customers and consumers with less benefit/engagement (e.g. those with low energy use) but prioritise customers where debt management is an issue, to reduce avoidable debt. This would distort results in relation to expectations for the full roll-out; therefore we need to understand whether the early mover installations are representative. Two approaches were identified for dealing with this:

- accept that lessons need to be learned from easier installations before moving on to the more difficult ones;
- expect some difficult installations to be made in the early Foundation phase, to ensure that practical issues are fully understood and resolved (e.g. communications solutions for flats) and differences in impact are identified.

Capabilities on project:  
Building Engineering

The latter could target all homes in a limited geographical area and may require Government funding support. The strategy needs to be decided as soon as possible.

Other points made were:

- representative sampling will simplify statistical analysis but segmentation choice will vary between suppliers;
- an upfront and honest approach to recruitment will generate a better response from consumers – to do otherwise runs the risk of the roll-out being halted because of privacy issues and campaigns.

### 8.5.3 Data collection

Two kinds of data collection were envisaged: routine data flows, and collecting data specifically to evaluate the roll-out. In both cases, a key concern was the massive storage, processing and retrieval challenge.

Routine data flows were seen as being limited in several ways.

- Evaluation requires good baseline data and this is limited. Where it depends on non-smart meter data (with relatively poor quality), larger numbers of homes and long periods will need to be used. Alternatively, where additional smart-enabled services are being evaluated, smart meter data could be used (but for relatively small numbers of homes in the Foundation stage). Suppliers may simply have to confirm what baseline they have used and state any caveats with the information.
- There were particular concerns over the early period (with an expectation that pre-DCC processes would be “more cumbersome” and have more “teething problems”) leading to customer service issues. Suppliers are building complex IS systems, and it will be challenging to get data out of these (if requirements are not specified early enough).
- The Annual Quantity database is useful for looking across regions but may not be very helpful for understanding impacts of individual interventions. Also, it will be important to extract these data at the right point in processing, e.g. before/after temperature correction.
- ELEXON's profile sampling meter base may be relevant, providing accurate, representative data. The timing of when a smart meter is installed will affect the apparent impact, so data on the timing of installations is required.
- If suppliers rely on customers opting in to have access to their detailed (half-hourly) data, this would introduce bias between customer segments as levels of opting in would vary.

In relation to dedicated data collection, a major theme was a need to be selective – it will not be feasible to measure everything and there is a trade-off between doing a few things well or many poorly (quality vs quantity). In addition, suppliers may already have determined clear trial objectives and data requirements, and it could be difficult to add to these. In any case, there could be customer resistance to collecting more data about them than suppliers do currently. However, it will be important to have the data required to pick up signals from the background noise, and distinguish minor and major effects.

Energy use is dependent on many factors that are not the main focus of suppliers' interests (although varying across suppliers), hence the necessary data may be lacking. Some limitations of suppliers' access to data were noted in Section 8.5.1. A further source of variation that needs to be controlled for is consistency of message in customer contacts. Contacts may go to different points within a supplier organisation (e.g. general call centres, call centres dedicated to smart meters, or the installers), resulting in different messages. Suppliers saw it as impossible to standardise supplier contacts for the purposes of evaluation.

Capabilities on project:  
Building Engineering

Other points made were as follows.

- Matching data across trials, samples and experimental/control groups will require timing data (e.g. in relation to effects of energy price changes, weather and cultural events such as August holidays and Christmas).
- While methods that are already established may appear to offer least cost, conversion from methods actually being used by suppliers may prove difficult and expensive.
- There may also be practical issues in terms of what level of detail is achievable during the Foundation stage.

#### 8.5.4 Customer feedback

This is a specific aspect of data collection and the general points from Section 8.5.3 therefore also apply here.

Consistency of methods was discussed in relation to the potential value of a short harmonised questionnaire to use in all data collection exercises (possibly by self-completion). Questions would always need to be asked in the same sequence, before other, non-standard questions. The standard questions would then influence responses to the non-standard questions (which might be the main focus of the suppliers' interest).

It also matters who administers a questionnaire, so there would need to be a standard type of interviewer (supplier, Government, University, etc.). This was seen as important in relation to the potential for bias in customer responses (e.g. over-estimating or over-reporting changes in behaviour) or willingness to provide personal household data at all. A market research organisation, independent of the industry and Government, may get the most accurate perspective. It was also felt that consumer groups can (and probably will) undertake these exercises anyway, the implication perhaps being that DECC and suppliers do not need to worry about it.

However, this then raises issues over linking datasets (e.g. because of data protection requirements) so that either the energy supplier has feedback data or the other organisation has consumption data, at household level. Data privacy was also seen more generally as a barrier, particularly if customers have the right to opt out of data linking (and there being a risk of this being the norm).

Other specific points made were:

- feedback is more important from consumers where smart meters have *not* benefited them;
- there may be customer fatigue with surveys and the number of interventions;
- there may be restrictions on suppliers' ability to talk to customers about energy-saving products at the point of smart meter installation and this will impact consumption reduction;
- there needs to be a plan for how (and how often) to capture relapse in behaviour.

#### 8.5.5 Data analysis

An important general point concerned who should carry out the analysis. This related to the conduct of the analysis but, more importantly, to independence and trust in the findings. Data collection/evaluation by an independent body or Government may be better received by consumers. Similarly, it is important to be clear about the audience for the analysis (e.g. suppliers, DECC, Ofgem, the public) and what data/analysis can be aggregated for public consumption or used internally (by suppliers or Government) for refinement of the Programme.

A particular challenge for the analysis will be multiple policies impacting on the same outcomes. The analysis needs to take account of the timing of other interventions (e.g. Green Deal installations). It will also need to take into account varying approaches to the Foundation stage (actions trialled in the Foundation stage could create bias later) and to monitor additional third party interventions. More generally, there is a moving baseline of customer knowledge and understanding: these are constantly changing and the analysis will need to control for this.

Capabilities on project:  
Building Engineering

## Annex A. Supplementary Documents

This report is supported by supplementary documents, supplied separately to DECC. They are referenced at the appropriate point in the report and, for convenience, are listed here.

<i>Document</i>	<i>Title</i>	<i>Content</i>
A	Workshop invitation	Letter of invitation sent to prospective workshop participants.
B	Background paper	Short background paper for participants, setting out the general aims of the workshop.
C	Workshop slides	All slides presented at the workshop.
D	Criteria list	List of criteria in the format presented to participants during the workshop (a simple list can be seen in Annex B).
E	Good practice posters	A3 posters of evaluation good practice points.

Capabilities on project:  
Building Engineering

## Annex B. Criteria for evaluation

1. The quality of *smart metering installation, services and support* in the roll-out as a whole.
2. The quality of *energy supply offerings* in the roll-out as a whole.
3. The quality of *information, feedback and support available to help consumers choose the most beneficial products and services* in the roll-out as a whole.
4. The quality of *data security and privacy* in the roll-out as a whole.
5. The degree of *competition in relevant markets* in the roll-out as a whole.
6. *Savings on domestic energy bills* in the roll-out as a whole.
7. Improvements in *customer convenience* in the roll-out as a whole.
8. Improvements in *consumers' energy awareness/knowledge* in the roll-out as a whole.
9. Changes in *consumers' energy-related behaviour* in the roll-out as a whole.
10. *The uptake of smart-enabled products and services (other than new tariffs)* in the roll-out as a whole.
11. *The uptake of new tariffs* in the roll-out as a whole.
12. Improvements in *energy supply efficiency* in the roll-out as a whole.
13. Improvements in *energy supply capacity* in the roll-out as a whole.
14. Improvements in *customer management efficiency (e.g. process simplification)* in the roll-out as a whole.
15. Improvements in *energy network efficiency* in the roll-out as a whole.
16. Improvements in *energy network capacity* in the roll-out as a whole.
17. Improvements in *energy generation efficiency* in the roll-out as a whole.
18. Improvements in *energy generation capacity* in the roll-out as a whole.
19. The *quality of consumer trials* in the roll-out as a whole.
20. *Reduced overall energy demand* in the roll-out as a whole.
21. *Smoother time profile of domestic energy use* in the roll-out as a whole.
22. Improvements in *industry cost-effectiveness* in the roll-out as a whole.
23. *Savings on domestic energy bills* in relation to specific products and services.
24. Improvements in *customer satisfaction* in relation to specific products and services.
25. Changes in *consumers' energy-related behaviour* in relation to specific products and services.
26. *Reduced overall energy demand* in relation to specific products and services.
27. *Smoother time profile of domestic energy use* in relation to specific products and services.
28. The quality of *consumer trials* in relation to specific products and services.
29. The degree of *competition in relevant markets* in relation to specific products and services.
30. Improvements in *energy supply efficiency* in relation to specific products and services.
31. Improvements in *energy supply capacity* in relation to specific products and services.
32. Improvements in *customer management efficiency (e.g. process simplification)* in relation to specific products and services.
33. Improvements in *industry cost-effectiveness* in relation to specific products and services.
34. Improvements in *consumers' energy awareness/knowledge* in relation to specific consumer engagement approaches.
35. *The uptake of smart-enabled products and services (other than new tariffs)* in relation to specific consumer engagement approaches.
36. *The uptake of new tariffs* in relation to specific consumer engagement approaches.

Capabilities on project:  
Building Engineering

## Annex C. Detailed findings on evaluation criteria

**Table C1** Ratings of importance of criteria – top half

Criterion	All participants		Suppliers		Others	
	Mean	SD	Mean	SD	Mean	SD
4. The quality of <i>data security and privacy</i> in the roll-out as a whole.	1.00	0.00	1.00	0.00	1.00	0.00
1. The quality of <i>smart metering installation, services and support</i> in the roll-out as a whole.	1.00	0.00	1.00	0.00	1.00	0.00
20. Reduced <i>overall energy demand</i> in the roll-out as a whole.	0.96	0.13	0.93	0.19	1.00	0.00
9. Changes in <i>consumers' energy-related behaviour</i> in the roll-out as a whole.	0.96	0.13	0.93	0.19	1.00	0.00
19. The <i>quality of consumer trials</i> in the roll-out as a whole.	0.93	0.27	0.86	0.38	1.00	0.00
3. The quality of <i>information, feedback and support available to help consumers choose the most beneficial products and services</i> in the roll-out as a whole.	0.93	0.27	1.00	0.00	0.86	0.38
26. Reduced <i>overall energy demand</i> in relation to specific products and services.	0.86	0.31	0.86	0.38	0.86	0.24
23. Savings on <i>domestic energy bills</i> in relation to specific products and services.	0.79	0.38	0.71	0.39	0.86	0.38
25. Changes in <i>consumers' energy-related behaviour</i> in relation to specific products and services.	0.79	0.38	0.71	0.39	0.86	0.38
2. The quality of <i>energy supply offerings</i> in the roll-out as a whole.	0.79	0.43	0.86	0.38	0.71	0.49
11. The <i>uptake of new tariffs</i> in the roll-out as a whole.	0.79	0.38	0.86	0.24	0.71	0.49
7. Improvements in <i>customer convenience</i> in the roll-out as a whole.	0.79	0.43	1.00	0.00	0.57	0.53
34. Improvements in <i>consumers' energy awareness/knowledge</i> in relation to specific consumer engagement approaches.	0.77	0.39	0.93	0.19	0.58	0.49
28. The <i>quality of consumer trials</i> in relation to specific products and services.	0.75	0.38	0.71	0.39	0.79	0.39
6. Savings on <i>domestic energy bills</i> in the roll-out as a whole.	0.75	0.43	0.64	0.48	0.86	0.38
14. Improvements in <i>customer management efficiency</i> (e.g. process simplification) in the roll-out as a whole.	0.75	0.38	0.86	0.24	0.64	0.48
24. Improvements in <i>customer satisfaction</i> in relation to specific products and services.	0.75	0.43	0.93	0.19	0.57	0.53
8. Improvements in <i>consumers' energy awareness/knowledge</i> in the roll-out as a whole.	0.71	0.47	0.86	0.38	0.57	0.53
5. The degree of <i>competition in relevant markets</i> in the roll-out as a whole.	0.71	0.38	0.71	0.39	0.71	0.39

Capabilities on project:  
Building Engineering

**Table C2** Ratings of importance of criteria – bottom half

Criterion	All participants		Suppliers		Others	
	Mean	SD	Mean	SD	Mean	SD
29. The degree of <i>competition in relevant markets</i> in relation to specific products and services.	0.65	0.38	0.79	0.27	0.50	0.45
33. Improvements in <i>industry cost-effectiveness</i> in relation to specific products and services.	0.64	0.46	0.71	0.49	0.57	0.45
12. Improvements in <i>energy supply efficiency</i> in the roll-out as a whole.	0.61	0.40	0.57	0.45	0.64	0.38
13. Improvements in <i>energy supply capacity</i> in the roll-out as a whole.	0.61	0.45	0.50	0.50	0.71	0.39
27. <i>Smoother time profile</i> of domestic energy use in relation to specific products and services.	0.58	0.49	0.43	0.53	0.75	0.42
35. The <i>uptake of smart-enabled products and services</i> (other than new tariffs) in relation to specific consumer engagement approaches.	0.58	0.49	0.71	0.49	0.42	0.49
36. The <i>uptake of new tariffs</i> in relation to specific consumer engagement approaches.	0.57	0.47	0.71	0.49	0.43	0.45
32. Improvements in <i>customer management efficiency</i> (e.g. process simplification) in relation to specific products and services.	0.54	0.48	0.64	0.48	0.42	0.49
10. The <i>uptake of smart-enabled products and services</i> (other than new tariffs) in the roll-out as a whole.	0.50	0.48	0.57	0.45	0.43	0.53
21. <i>Smoother time profile</i> of domestic energy use in the roll-out as a whole.	0.46	0.46	0.43	0.45	0.50	0.50
31. Improvements in <i>energy supply capacity</i> in relation to specific products and services.	0.46	0.48	0.36	0.48	0.58	0.49
30. Improvements in <i>energy supply efficiency</i> in relation to specific products and services.	0.46	0.48	0.64	0.48	0.25	0.42
22. Improvements in <i>industry cost-effectiveness</i> in the roll-out as a whole.	0.43	0.43	0.50	0.50	0.36	0.38
18. Improvements in <i>energy generation capacity</i> in the roll-out as a whole.	0.36	0.41	0.43	0.45	0.29	0.39
16. Improvements in <i>energy network capacity</i> in the roll-out as a whole.	0.36	0.41	0.21	0.39	0.50	0.41
15. Improvements in <i>energy network efficiency</i> in the roll-out as a whole.	0.32	0.42	0.07	0.19	0.57	0.45
17. Improvements in <i>energy generation efficiency</i> in the roll-out as a whole.	0.29	0.38	0.43	0.45	0.14	0.24

Capabilities on project:  
Building Engineering

**Table C3** Ratings of PAC – top half

Criterion	All participants		Suppliers		Others	
	Mean	SD	Mean	SD	Mean	SD
20. Reduced <i>overall energy demand</i> in the roll-out as a whole.	0.93	0.27	1.00	0.00	0.86	0.38
23. Savings on <i>domestic energy bills</i> in relation to specific products and services.	0.89	0.29	1.00	0.00	0.79	0.39
19. The <i>quality of consumer trials</i> in the roll-out as a whole.	0.82	0.37	0.79	0.39	0.86	0.38
26. Reduced <i>overall energy demand</i> in relation to specific products and services.	0.82	0.32	0.79	0.39	0.86	0.24
28. The <i>quality of consumer trials</i> in relation to specific products and services.	0.82	0.32	0.86	0.24	0.79	0.39
4. The quality of <i>data security and privacy</i> in the roll-out as a whole.	0.79	0.38	1.00	0.00	0.57	0.45
6. Savings on <i>domestic energy bills</i> in the roll-out as a whole.	0.79	0.43	1.00	0.00	0.57	0.53
9. Changes in <i>consumers' energy-related behaviour</i> in the roll-out as a whole.	0.75	0.38	0.71	0.39	0.79	0.39
25. Changes in <i>consumers' energy-related behaviour</i> in relation to specific products and services.	0.75	0.38	0.79	0.39	0.71	0.39
33. Improvements in <i>industry cost-effectiveness</i> in relation to specific products and services.	0.75	0.38	0.93	0.19	0.57	0.45
31. Improvements in <i>energy supply capacity</i> in relation to specific products and services.	0.73	0.39	0.86	0.38	0.58	0.38
1. The quality of <i>smart metering installation, services and support</i> in the roll-out as a whole.	0.71	0.38	0.93	0.19	0.50	0.41
27. <i>Smoother time profile</i> of domestic energy use in relation to specific products and services.	0.69	0.43	0.50	0.50	0.92	0.20
30. Improvements in <i>energy supply efficiency</i> in relation to specific products and services.	0.69	0.43	0.79	0.39	0.58	0.49
34. Improvements in <i>consumers' energy awareness/knowledge</i> in relation to specific consumer engagement approaches.	0.65	0.47	0.79	0.39	0.50	0.55
14. Improvements in <i>customer management efficiency</i> (e.g. process simplification) in the roll-out as a whole.	0.64	0.41	0.93	0.19	0.36	0.38
36. The <i>uptake of new tariffs</i> in relation to specific consumer engagement approaches.	0.64	0.41	0.71	0.39	0.57	0.45
21. <i>Smoother time profile</i> of domestic energy use in the roll-out as a whole.	0.64	0.46	0.71	0.49	0.57	0.45

Capabilities on project:  
Building Engineering

**Table C4** Ratings of PAC – bottom half

Criterion	All participants		Suppliers		Others	
	Mean	SD	Mean	SD	Mean	SD
3. The quality of <i>information, feedback and support available to help consumers choose the most beneficial products and services</i> in the roll-out as a whole.	0.61	0.45	0.79	0.39	0.43	0.45
2. The quality of <i>energy supply offerings</i> in the roll-out as a whole.	0.57	0.47	0.64	0.48	0.50	0.50
22. Improvements in <i>industry cost-effectiveness</i> in the roll-out as a whole.	0.57	0.43	0.79	0.39	0.36	0.38
12. Improvements in <i>energy supply efficiency</i> in the roll-out as a whole.	0.54	0.41	0.64	0.38	0.43	0.45
11. The <i>uptake of new tariffs</i> in the roll-out as a whole.	0.50	0.48	0.64	0.48	0.36	0.48
35. The <i>uptake of smart-enabled products and services</i> (other than new tariffs) in relation to specific consumer engagement approaches.	0.50	0.46	0.57	0.45	0.42	0.49
32. Improvements in <i>customer management efficiency</i> (e.g. process simplification) in relation to specific products and services.	0.50	0.50	0.71	0.49	0.25	0.42
10. The <i>uptake of smart-enabled products and services</i> (other than new tariffs) in the roll-out as a whole.	0.50	0.48	0.57	0.45	0.43	0.53
18. Improvements in <i>energy generation capacity</i> in the roll-out as a whole.	0.46	0.41	0.57	0.45	0.36	0.38
24. Improvements in <i>customer satisfaction</i> in relation to specific products and services.	0.39	0.49	0.57	0.53	0.21	0.39
8. Improvements in <i>consumers' energy awareness/knowledge</i> in the roll-out as a whole.	0.39	0.45	0.50	0.50	0.29	0.39
17. Improvements in <i>energy generation efficiency</i> in the roll-out as a whole.	0.39	0.40	0.43	0.45	0.36	0.38
7. Improvements in <i>customer convenience</i> in the roll-out as a whole.	0.36	0.46	0.57	0.53	0.14	0.24
13. Improvements in <i>energy supply capacity</i> in the roll-out as a whole.	0.36	0.41	0.43	0.45	0.29	0.39
16. Improvements in <i>energy network capacity</i> in the roll-out as a whole.	0.36	0.41	0.21	0.39	0.50	0.41
5. The degree of <i>competition in relevant markets</i> in the roll-out as a whole.	0.32	0.37	0.36	0.38	0.29	0.39
15. Improvements in <i>energy network efficiency</i> in the roll-out as a whole.	0.29	0.38	0.36	0.48	0.21	0.27
29. The degree of <i>competition in relevant markets</i> in relation to specific products and services.	0.23	0.33	0.36	0.38	0.08	0.20

Capabilities on project:  
Building Engineering

**Table C5** Rank order of differences in rating between suppliers and other participants

<i>Importance</i>		<i>PAC</i>	
<i>Criterion</i>	<i>Difference</i>	<i>Criterion</i>	<i>Difference</i>
7	0.43	14	0.57
30	0.39	32	0.46
24	0.36	1	0.43
34	0.35	4	0.43
35	0.30	6	0.43
36	0.29	7	0.43
8	0.29	22	0.43
17	0.29	33	0.36
29	0.29	3	0.36
32	0.23	24	0.36
14	0.21	11	0.29
3	0.14	34	0.29
33	0.14	29	0.27
10	0.14	31	0.27
18	0.14	12	0.21
22	0.14	8	0.21
2	0.14	23	0.21
11	0.14	18	0.21
1	0.00	30	0.20
4	0.00	35	0.15
5	0.00	36	0.14
26	0.00	2	0.14
9	-0.07	20	0.14
20	-0.07	21	0.14
28	-0.07	15	0.14
21	-0.07	10	0.14
12	-0.07	13	0.14
23	-0.14	5	0.07
25	-0.14	17	0.07
19	-0.14	28	0.07
6	-0.21	25	0.07
13	-0.21	26	-0.07
31	-0.23	9	-0.07
16	-0.29	19	-0.07
27	-0.32	16	-0.29
15	-0.50	27	-0.42

A positive value means the suppliers gave the higher ratings.

Capabilities on project:  
Building Engineering

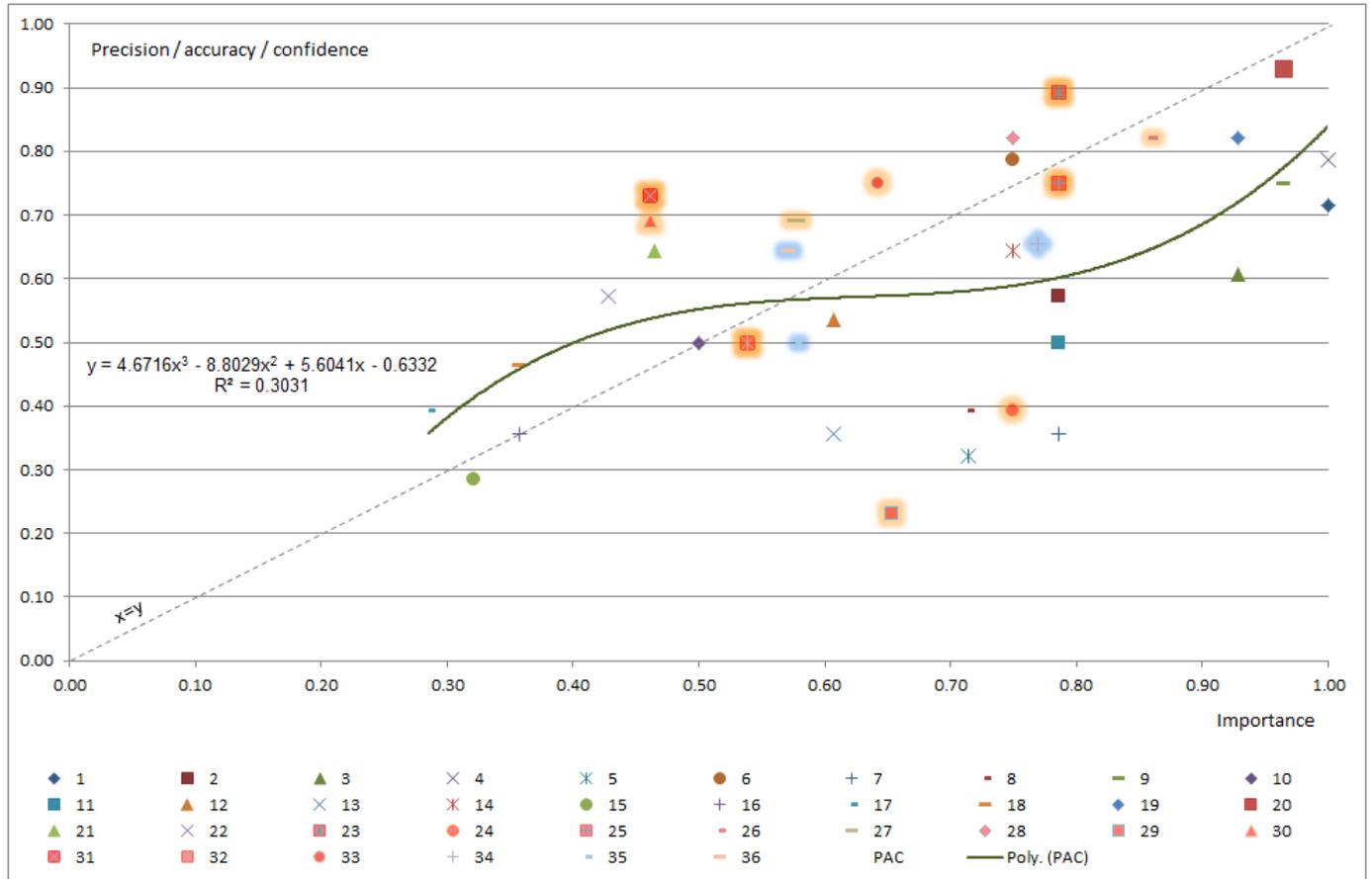
**Figure C1** Table placements of criteria



Table A placements are shown in red, Table B in blue and Table C in green.

Capabilities on project:  
Building Engineering

**Figure C2** Plot of PAC against importance ratings



Capabilities on project:  
Building Engineering

**Figure C3** Plot of “other” participant against supplier ratings

