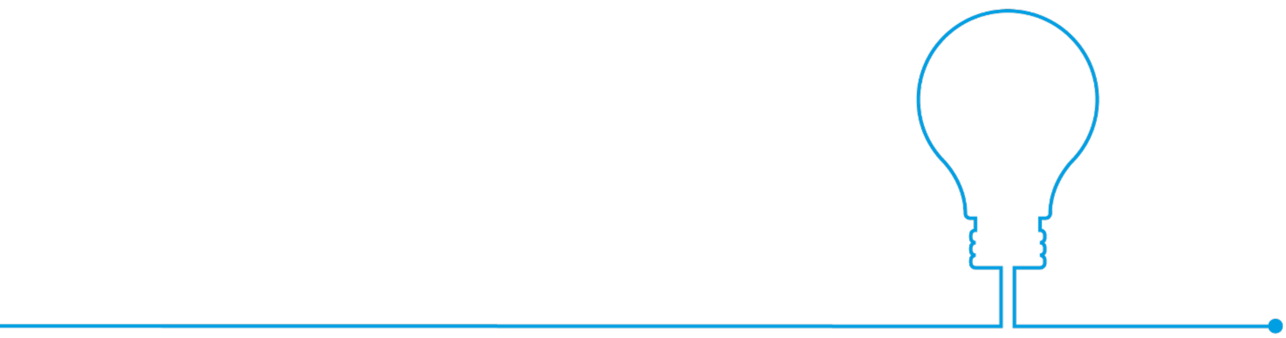
Department for Business, Energy & Industrial Strategy

EVALUATION OF THE TRANSITIONAL ARRANGEMENTS FOR DEMAND SIDE RESPONSE

Phase 2 – Main Report



December 2017

Contents

1 Introduction

This report presents findings from the evaluation of the first auction of the Transitional Arrangements (TA) for Demand Side Response (DSR) and small-scale distribution-connected generation. This realist, theory-based evaluation was undertaken for the Department for Business, Energy and Industrial Strategy (BEIS) by CAG Consultants, in partnership with Databuild, Verco and NERA Economic Consulting. An earlier report[1] from Phase 1 of the evaluation presented findings about the auction, including awareness of the auction and pre-qualification, while this Phase 2 report focuses on delivery of the TA obligations after the auction.

Research and policy background

The TA is a pilot and forms part of the Capacity Market (CM) for security of electricity supply. The TA aims to support BEIS's objectives of promoting growth, decarbonisation and energy security, while ensuring affordability of the energy supply.

The TA aims to encourage development of DSR or small-scale distribution-connected generation that is increasingly needed to balance supply and demand in a decarbonised electricity grid.[2] In this report we use the CM definition of DSR: the activity of reducing the metered volume of imported electricity of one or more customers below an established baseline, by means other than a permanent reduction in electricity use. Under this definition, DSR may be achieved through any combination of onsite generation, temporary demand reduction or load-shifting. We use the term 'turn-down' DSR to refer to the last two activities.

The TA scheme involved two auctions for specific types of capacity within the CM, the first for delivery of capacity in the 2016/17 delivery year[3] and the second for delivery of capacity in 2017/18. These TA auctions are additional to the main CM auctions: the main four-year ahead auctions (T-4) and the smaller one-year ahead auctions (T-1) which will deliver capacity from 2018/19 onwards, as well as the Early Auction which BEIS held in early 2017 to deliver capacity in 2017/18.

[1] https://www.gov.uk/government/publications/evaluation-of-the-transitional-arrangements-phase-1
[2] National Infrastructure Commission (2016) *Smart Power: A National Infrastructure Commission Report.* Available at: https://www.gov.uk/government/publications/smart-power-a-national-infrastructure-commission-report. Accessed 27/7/2016
[3] The delivery year runs from 1st October of one year through to 30th September of the following year.

The TA has three main objectives, which we have used as the basis for three project hypotheses to be tested by the evaluation (see below and Appendix 1 for further details):

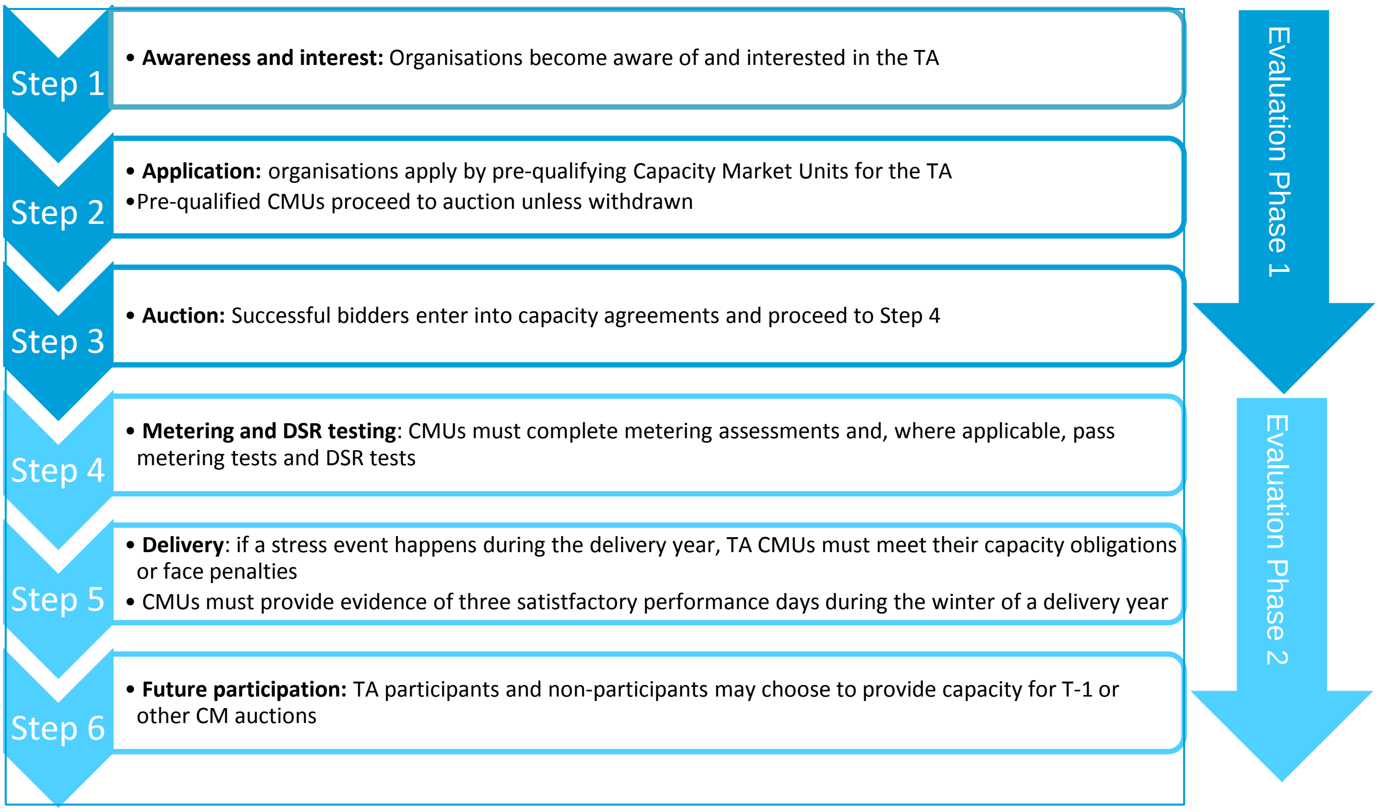
1. To contribute to security of electricity supply to help with short-term forecasted system tightness (winter 2016/17 and winter 2017/18).

2. To develop a stock of flexible capacity[4] that can be available for the one-year ahead (T-1) auction in 2017 for delivery in 2018/19, thereby contributing to liquidity in this and subsequent year-ahead auctions.

3. To encourage enterprise and develop experience, confidence and understanding so that DSR and embedded generation will be able to realise their potential and ultimately compete with larger generation assets in the CM.

The diagram below sets out the six steps in the TA process. The Phase 1 report covered steps 1-3 for the first TA auction, while this Phase 2 report focuses primarily on steps 4-6.

Figure 1.1 Steps in the TA process



[4] Flexible capacity means electricity generating capacity and demand that is able to increase or decrease in response to signals, to help balance supply and demand of electricity across the GB grid.

The first TA, which is evaluated here, was open to providers of DSR and small-scale distribution-connected generation services. The second TA auction was open to 'turn-down' DSR only. Phases 3 and 4 of the evaluation will focus on the second TA.

Evaluation aims and objectives

The high-level evaluation questions (HLQs) specified by BEIS for this evaluation are shown in Figure 1.2. The evaluation responds to these questions.

Figure 1.2 High level evaluation questions

High level question
1. What outcomes can be attributed to the TA and were they as intended by BEIS? What outcomes occurred for whom and under what circumstances?
2. Through what levers and causal mechanisms has the TA contributed to these outcomes and the variation by group and circumstance?
3. Did the TA represent good value for money to both scheme participants and the consumer?
4. Which aspects of the TA's design and implementation account for the findings of HLQ 2 and 3?
5. What are the implications of the findings for the future contribution of DSR and small-scale generation to the CM?

Evaluation design

Our approach to this evaluation is realist and theory-based. A realist approach[5] emphasises the importance of understanding not only whether a policy contributes to outcomes (which may be intended or unintended) but how, for whom and in what circumstances. We chose a realist approach for this evaluation because the number of organisations involved in the TA is small (precluding statistical approaches to assessing the policy's impact on outcomes) and the policy area is complex (requiring careful unpicking of 'how' and 'why' organisations made the choices they did). The emphasis of the realist approach on understanding 'how' and 'why' the TA influenced – or did not influence - certain types of organisations will help the evaluation to inform the roll out of DSR and small-scale generation within the main CM.

[5] Pawson and Tilley (1997) (op cit); Pawson (2006) (op cit).

The development of a ‘theory’ of the TA is central to implementing a realist evaluation as it allows evaluators to rigorously examine the design and execution of the scheme, and test policy assumptions against available evidence. We developed a theoretical framework for Phase 2 of the evaluation, as presented in Appendix 2, which sets out the realist hypotheses that we tested against research evidence. The realist hypotheses set out for whom, and in what circumstances (i.e. in what ‘contexts’), the policy is expected to lead to particular reasoning and choices being made (i.e. causal ‘mechanisms’ being activated[6]), leading to desired or undesired policy outcomes. These realist hypotheses are generally known as context-mechanism-outcome combinations or ‘CMOs’[7].

Realist evaluation uses the idea of generative causality (i.e. a mechanism or reasoning only fires when the contexts are right). We therefore used generative causation assessment methods which involve a forensic examination of causality using strategic data collection and logic, rather than a probabilistic assessment of causation through statistical correlation[8]. This allowed hypotheses about causality and impact to be developed and tested at the participant and CMU level using in-depth information about the influence of the TA and alternative drivers for change.

The main generative causation method that we used in Phase 2 of the evaluation was realist contribution analysis involving the formulation of CMOs to explain the behaviour of TA participants, and their subsequent revision based on careful assessment of observed evidence from interviews, public statements, analysis of scheme data and case studies of testing data. This process, summarised in Figure 1.3 below, was informed by contribution tracing with Bayesian updating. Contribution tracing is a formal process involving the specification of the types of evidence that would be observed to support the TA hypotheses or competing causal hypotheses, and testing of the strength of these hypotheses against the evidence actually observed.

[6] In realist terminology, the activation of a causal mechanism is referred to as the mechanism ‘firing’.
[7] Definitions for contexts, mechanisms and outcomes are provided in the glossary. Further detail can be found in Pawson and Tilley (1997) (op cit).
[8] Because of the small sample sizes, we did not just look for statistical correlations in the evaluation evidence (e.g. x% of a given type of organisation demonstrated a particular outcome) but aimed to understand how and why organisations made the choices they did, on a case by case basis.

Figure 1.3 Summary of the realist contribution analysis process

These methods were used to analyse what mechanisms (reasoning) led to particular outcomes (behaviour) for TA participants during steps 4-6 of the TA process, and to identify the contexts (attributes, resources, knowledge, skills, beliefs, etc.) that, in combination with the TA, triggered those mechanisms. The generative causation methods are explained further in Appendix 4 and 5, while the revised theoretical framework is presented in Appendix 3.

Methodology

Research undertaken in Phases 1 and 2 of the evaluation has tested the initial theoretical framework against the sources of evidence listed in Figure 1.4 below.

The topic guides and email questionnaire for Phase 2 research were agreed in advance with BEIS. They were designed to gather the evidence required to test the theoretical framework, to apply the contribution tracing tests and to fill gaps in Phase 1 evidence on the characteristics and costs of TA capacity. Interviews were recorded and Phase 2 evidence for each participant was analysed in spreadsheet format, supplemented by Phase 1 evidence where required. This evidence was analysed against the hypothesised CMOs in the theoretical framework. Revised CMOs were developed on a case by case basis for each element of the framework. Further detail on the methodology for qualitative research is provided in Appendices 6, 7 and 8, while our approach to analysis is set out in Appendix 4.

The capacity provided through the first TA was characterised using data on capacity and testing outcomes from BEIS and National Grid, combined with capacity estimates provided in Phase 2 interviews, and supplemented by the Phase 2 email survey and Phase 1 interview findings where appropriate. We used this evidence to estimate the proportion of capacity in each CMU that was dependent on TA revenue, the proportion that was provided by turn-down DSR, the proportion provided by baseload generation and to analyse how different elements of TA capacity responded to Capacity Market Notices (CMNs). This allowed refinement of the capacity estimates made during Phase 1 of the evaluation.

Contribution tracing analysis with Bayesian updating was undertaken on a case by case basis during Phase 2, using pre-specified evidence tests based on both observed behaviour and interview statements, drawing on evidence from both Phase 1 and Phase 2. The findings of contribution tracing were used to cross check the CMO analysis for each case (e.g. whether and how each case contributed to each of the TA objectives) and the characterisation of TA capacity (e.g. the proportion of capacity that was dependent on TA revenue). Further information on the contribution tracing tests is provided in Appendix 5.

In addition, case studies of DSR and Satisfactory Performance Day (SPD) tests were undertaken during Phase 2 to assess the reliability of DSR, by analysing the variability of the baseline demand against which demand reductions were measured. These case studies involved analysis of meter data from DSR and SPD tests for five CMUs with turn-down DSR components, put forward by four different participants. Further details of the case studies are provided in Appendix 9.

The research summarised in Figure 1.4 involved research with direct participants that put their own capacity into the TA, and research with aggregators[9] who put forward capacity on behalf of other organisations.

Figure 1.4 Evidence sources for Phases 1 and 2 of the evaluation

Evidence source	Phase 1	Phase 2
Analysis of data from CM register	✓	✓
Analysis of auction behaviour	Analysis of TA auction behaviour, including exit prices	High-level outcomes for DSR in Early Auction and T-4 auction.
Modelling of supply curve for auction	✓	-
Email survey of TA participants	-	17 out of 24 responses
In-depth interviews with TA participants	Interviews with all 24 participants in the TA auction, and with the 5 organisations that pre-qualified but did not	Interviews with 19 participants, including those exiting from the TA, supplemented by email responses from 4 participants

[9] An aggregator is an intermediary organisation that provides a service of collating capacity for flexibility services from a range of other organisations, in return for a share in the revenues generated.

Evidence source	Phase 1	Phase 2
	obtain capacity agreements.	(total 23 of the 24 participants).
In-depth interviews with clients of aggregators	4 interviews, identified via TA applications and non-participant survey	7 interviews with aggregator clients, identified via aggregators
Research with non-participants	31 interviews with organisations with potential for TA, plus a survey of 169 potential participants	-
Case study analysis	-	4 case studies of DSR and SPD tests to analyse baseline issue
In-depth interviews with policy makers and delivery bodies	✓	✓
Workshops to test emerging findings	Internal workshop with BEIS and peer reviewers; external workshop with delivery bodies and industry representatives	Internal workshop with BEIS and peer reviewers.

Further details on the research and analysis undertaken during Phase 2, and our rationale for Phase 2 data collection, are presented in Appendices 4-9.

Limitations of this research

The findings of this Phase 2 research are robust, being based on extensive research with TA participants, on a wide range of evidence sources, and on realist contribution analysis which has been cross-checked using rigorous contribution tracing methods. Nevertheless, the research has a few limitations:

- There has not yet been a stress event in the current delivery year so there is little evidence to date about delivery of TA obligations. We have based our analysis on participant's self-reported responses to the two Capacity Market Notices (CMNs), as no meter data is available in the absence of a stress event. While further CMNs or stress events could happen before the end of the delivery year in September 2017, this is unlikely to occur during the summer when electricity demand is low.

- Identifying aggregator clients has been problematic, as the delivery bodies hold meter point details but not organisation names. Some aggregators were willing to share the identity of their clients, but the sample of aggregator clients is limited.
- While we can observe TA organisations' participation in other CM auctions, we have little information on bid prices as these are confidential. During this research, we had no access to bid prices from the T-4, Early Auction or second TA auction, which limited our ability to assess value-for-money by analysing the impact (if any) of participation in the first TA on bidding behaviour in other CM auctions.
- Phase 2 interviews generated information on the costs of meeting TA testing requirements, but little information on the opportunity cost for turn-down DSR.
- Our development of CMOs was limited in some cases by lack of depth in the evidence gathered during Phase 2. Seven participants did not complete the email survey and four offered email responses in place of an in-depth interview but there was only one participant from whom we had no input during Phase 2. In-depth interview respondents did not always have insights into the full range of organisational experiences on the TA (e.g. because they were new in post or were not close to some details of the organisation's operations). But in all cases we had sufficient data to develop some form of CMO.

Report outline

- Section 2 - findings on testing (step 4 of Figure 1.1), and its implications for TA capacity.
- Section 3 - findings on delivery (step 5 of Figure 1.1), from available evidence.
- Sections 4, 5 & 6 - findings on the TA's contribution to its three objectives.
- Section 7 – recap on value-for-money from the TA.
- Section 8 - wider learning emerging from the evaluation, in relation to the CM and DSR.

2 Participation and testing in the first TA

Testing requirements

Following the award of capacity agreements for the first TA auction, participants had to pass standard CM tests to confirm their capacity. These tests are designed to ensure that the full capacity procured in CM auctions can be relied upon when needed. The tests are summarised in Figure 2.1 and explained further in Appendix 1.

Figure 2.1 Summary of Capacity Market testing requirements

Metering assessments are required for all Capacity Market Units (CMUs) to determine which metering option applies to each of their sites. Three metering options qualify, as follows:

(a) Supplier settlement metering;
(b) Bespoke metering; and
(c) Balancing services metering

Metering tests are required for sites using metering options (b) and (c), but not option (a).

DSR tests are required for unproven DSR CMUs to demonstrate that they can deliver the required demand reduction against a measured baseline of demand. The 'proven' capacity of the CMU reflects the outcome of DSR testing.

Three 'satisfactory performance days' (SPDs) are required for all CMUs to demonstrate that their capacity remains available through the winter delivery period.[10]

We proposed an initial theory as to how TA participants would react and respond to these testing requirements at the start of the evaluation, as shown in Appendix 2. To continue in the TA, participants had to pass all necessary tests, and so testing had implications for continued TA participation. Our revised theory for testing, taking into account the data collected during Phase 2, is shown in Appendix 3.

Summary of testing and participation outcomes for the first TA

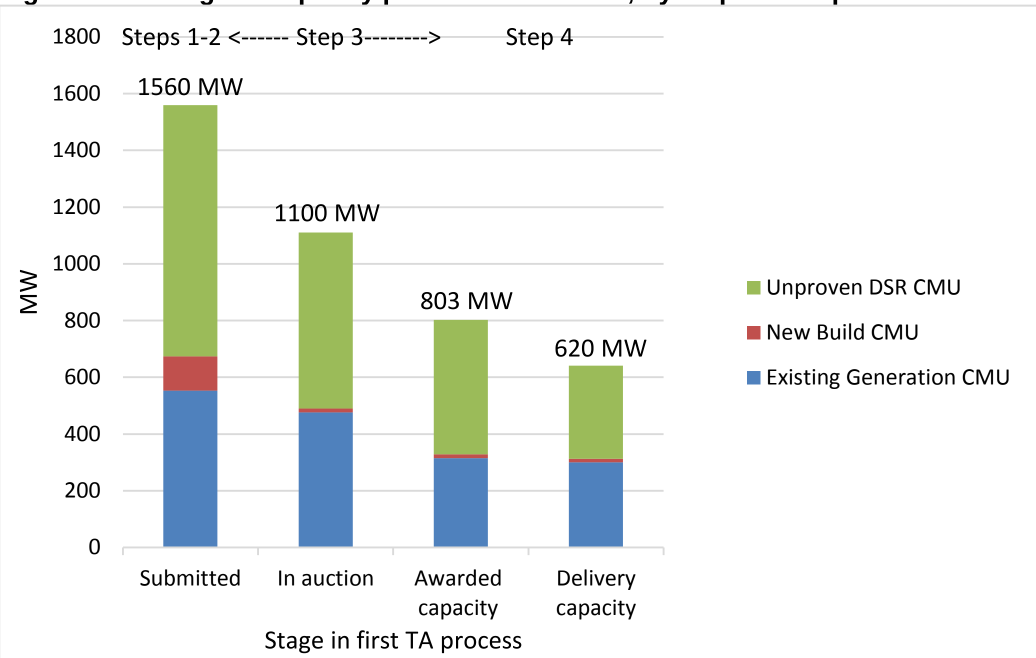
Around 23% of capacity, equal to 182.4MW dropped out of the TA due to the testing process. While 802.7 MW of capacity was awarded capacity agreements in the first TA

[10] The winter delivery period was defined as 1st October 2016 to 30th April 2017.

auction, only 620.3 MW of capacity has gone forward to delivery, as shown in Figure 2.2. Almost all of the lost capacity (166.8 MW) was classified as ‘unproven DSR’, representing around 35% reduction in DSR capacity and leaving 307.9 MW of ‘proven’ DSR in the TA, down from the 474.7 MW that was successful at auction. Only 15.5 MW of existing generation capacity dropped out after the action, representing around 5% of this type of capacity and leaving 299.7 MW of existing generation in the TA. A small amount of new build capacity (12.8 MW) passed the testing requirements without losing and capacity. The 620.3 MW includes 10.7 MW that is currently suspended at the time of writing for failing to pass SPDs but may pass on resubmission of SPD data.

While the stringency of the testing process was the primary reason behind loss of capacity, the testing process also revealed some capacity that was lost for other reasons (e.g. aggregators having problems signing up clients that could meet TA requirements). The underlying reasons for the loss of capacity are analysed in the next section below.

Figure 2.2 Change in capacity provided in first TA, by steps in TA process



Source: TA register

After testing, 46 CMUs went forward for delivery, comprising 620.3 MW across 21 organisations. In total, three organisations dropped out from the 24 that succeeded at auction (covering four CMUs of both unproven DSR and existing generation with a total capacity of 70.0 MW); seven unproven DSR CMUs (with a capacity of 51.4 MW) were withdrawn or terminated by organisations that still had other CMUs going forward to delivery; and a further seven unproven DSR CMUs went forward with reduced capacity (leading to a further loss of 60.9 MW). 11 organisations and a total of 39 CMUs passed testing without reducing capacity. These figures are summarised in Table 2.3.

Almost all (91%) of the lost capacity was unproven DSR, for reasons discussed below. Participants with existing generation CMUs did not have the option of passing with reduced capacity as their capacity was already 'proven'. Of the 299.7 MW of existing generation proceeding to delivery, around 90 MW initially failed the tests and passed on appeal (comprising all the existing generation that appealed). Of the 307.9 MW of proven DSR proceeding to delivery around 10 MW passed on appeal, while around 35 MW of the 166.8 MW of the lost DSR capacity appealed but failed to meet the appeal criteria.

Table 2.3 Breakdown of capacity lost post-auction

Category	MW	% DSR	Number of CMUs	Number of organisations
(a) CMU withdrawn or terminated, leaving organisation with no other CMUs in TA	70.0	78%	4	3
(b) CMU withdrawn or terminated, leaving organisation with other CMUs still in TA	51.4	100%	7	3
(c) capacity lost from CMUs that are going forward for delivery in the TA	60.9	100%	7	5
Total lost capacity (a + b + c)	**182.4**	**91%**	**18**	**10**
Capacity remaining in the TA	620.3	50%	46	21

Note: Total MWs do not add owing to rounding. Also, the number of organisations does not add to the total across the three categories because some organisations had CMUs in both categories (b) and (c).

We estimate that 70% of the capacity remaining in the TA was put forward by aggregators or potential aggregators, who put forward capacity belonging to other organisations or were considering doing so in future. Based on interview evidence, we identified 11 aggregators or potential aggregators remaining in the first TA, offering 431 MW across 33 CMUs. Ten direct participants put forward the remaining 189 MW, across 13 CMUs.

How, why and for whom these outcomes arose

Our realist analysis used the interview evidence and testing data to explore participants' reasoning about their situation and TA requirements. This analysis is summarised below, and gives the outcome that arose, the reasoning that led to the outcome and the contexts that resulted in the reasoning. One participant is omitted from this analysis owing to lack of Phase 2 data. Reasoning associated with positive outcomes are presented first.

Reasoning: 'TA testing requirements are as we anticipated, we'll just get on with it'

Five participants (representing 6 CMUs and nearly 100 MW) found TA testing requirements fairly straightforward to meet for their CMUs, in line with their expectations.

In all cases, these CMUs had relatively simple metering test requirements because they used metering option (a) (and hence required no metering test).

> *"When you're doing the metering assessment, there are basically two roads you can go down. You can go down the easy road if you are using a settlement meter, but as soon as you use any sort of complicated metering arrangement you have to go down the other road, and that road is unnecessarily complicated." (Aggregator)*

They also had little dependence on actions by third parties as most were direct participants with access to their own supplier settlement data. Other contexts that contributed to this reasoning in some cases were that:

- Four participants had relatively simple CMUs (e.g. one CMU, one component).
- Two participants offering DSR could fit DSR tests and Satisfactory Performance Days (SPDs) around the needs of an individual site, or already had access to meter data to facilitate these tests.
- Three participants were offering small-scale generation that did not require DSR tests and two were running as baseload, meaning they were delivering anyway for SPDs.
- Those offering small-scale generation were already set up to export to the grid (e.g. they were already being used to generate for Triad[11], STOR or FFR[12]).
- They had a knowledgeable 'committed individual' within the organisation or had contracted another organisation to provide this expertise.
- Some had prior knowledge and experience of flexibility markets.

Reasoning: 'TA testing requirements are more demanding/expensive than we anticipated, but it's still worthwhile'

For most CMUs, participants found some elements of the testing requirements (metering testing, DSR testing and SPDs) more demanding or more expensive than they had

[11] Transmission charges are based on electricity demand during three peak demand periods in a given year (called 'triads'). Many large consumers and generators try to reduce demand or increase generation at times that have a high risk of being the 'triads', to reduce their transmission charges.

[12] The EMR delivery body buys a number of flexibility services from electricity consumers and generators, including the Short Term Operating Reserve (STOR) and Fast Frequency Response (FFR).

anticipated, but still thought it worthwhile to continue with their full capacity in the TA. This reasoning was observed for 15 participants in relation to 32 CMUs (around 430-440 MW).

Six aggregators and one direct participant, together offering 16 unproven DSR CMUs (190-200 MW), reported that testing generated a high administration and cost burden (e.g. clarifying requirements with the delivery and settlement bodies; finding clients and sites that could meet metering requirements; investing in metering; and coordinating DSR tests and SPD delivery across many sites) but reasoned that it was still worthwhile to maintain their full capacity to retain TA revenues and to progress their strategic priorities for future CM participation. These aggregators put significant effort into recruiting clients for 'unfilled' CMUs, were well-prepared for the requirements of the TA and largely selected sites that could use metering option (a), in order to avoid metering tests.

Multiple sites offered more flexibility (e.g. if one site failed to deliver then other sites might be able to compensate by offering more capacity), but constrained aggregators' ability to fit DSR tests and SPDs with their clients' operational requirements, baseline situation and delivery for Triad or other services. This led to some tests being repeated or appealed. Prior knowledge and experience helped some participants minimise metering test 'pitfalls'. For example, there were examples of aggregators using their knowledge and T-4 experience to select suitable assets for their CMUs (e.g. avoiding assets with renewable generation on site) that would avoid them having to select more complex metering options (i.e. options (b) and (c)).

"Knowing we were going to have to use supplier settlement metering, and wanting to make sure we didn't have any components that didn't fit squarely within the easiest path to market, we didn't accept anybody that had on-site renewables that would require a special metering process." (Aggregator)

Nine aggregators and direct participants offering 16 'existing generation' or 'new build' CMUs (240-250 MW), including many sites that required metering options (b) or (c) (e.g. because there were onsite renewables which needed to be separately metered), put in significant effort to meet metering test requirements because they wanted to retain TA revenues (and also position themselves for future CM participation).

As 'proven' CMUs, they would have had to exit the TA if their capacity had been reduced. Examples of this effort included clarifying requirements with National Grid and EMR

Settlements (EMRS), obtaining historic documentation of metering accuracy, making major investments[13] in new equipment and/or submitting appeals.

> *"That [investment in metering] was a significant cost, but when that's weighed up against the benefits of the first year of TA alone then it's clear that it was still worthwhile for us to do it." (Direct participant)*

Reasoning: 'TA testing requirements are more demanding/expensive than we anticipated – and we had to drop some capacity due to factors beyond our control – but it's still worthwhile for the remaining capacity in this CMU.'

Five aggregators or potential aggregators dropped 61 MW of capacity across seven unproven DSR CMUs. These CMUs remained above the minimum threshold of 2 MW and these aggregators still regarded it as worthwhile to meet the testing requirements for the remainder of their capacity in these CMUs.

About half of the capacity reduction arose from aggregators who had put forward significant volumes of 'unfilled' unproven DSR CMUs and had problems signing up new clients that could pass testing requirements within the TA timeframe. This was particularly problematic for potential clients new to flexibility, where the process of getting these clients on board took longer than some anticipated. Those with sign-up problems reported that they would have done things differently if they had been better prepared. For example, some would have prioritised sites with settlement metering if they had realised the implications earlier.

The remaining capacity reduction arose from existing or potential aggregators who had signed up sufficient volumes of clients for their unproven DSR CMUs, or were testing the aggregation model using their own sites, but who failed tests for some components because of external factors beyond their control and did not have time for (or did not want to risk[14]) a retest. One-off failures included a back-up generator failing or a client failing to deliver during a test. Other problems included electricity suppliers being slow to set up data flows for supplier settlement meters, or new grid connections requiring action by the Distribution Network Operator (DNO). Reliance on external agencies was particularly problematic given the tight timeframe set by the TA.

[13] While most metering investments were reported to be a few thousand pounds, in a few cases costs were much higher (e.g. £50,000 - £120,000), either because investment was required for multiple sites or because replacement of major equipment (e.g. current transformers) was needed to achieve the required accuracy. Additional costs were associated with site closure to allow replacement of equipment.

[14] The perceived risk of a retest was that other component(s) might fail to deliver.

Reasoning: ‘The TA is still worthwhile but we could not meet the testing or eligibility requirements’

Four participants, across nine CMUs, withdrew or failed to meet testing or eligibility requirements for a whole CMU but indicated in interview that they would have continued to participate if they could. This included two of the three participants with no capacity left after testing. These participants’ circumstances were similar to those for the ‘reduced capacity’ group, except that they were unable to meet the minimum capacity for their CMUs to continue in the TA. For example:

- They were less well prepared for the TA and some had misunderstood meter requirements.
- There were external factors affecting their ability to meet requirements (e.g. reliance on DNO connections, or reliance on suppliers to set up metering data flows).
- Some of the sites required metering option (b) (e.g. because of renewable energy elements), which was not straightforward to implement.
- Aggregators faced problems signing up new clients to fill unproven DSR CMUs, particularly those clients that were new to flexibility.

Reasoning: ‘The TA is more demanding than we expected and it’s no longer worthwhile to participate directly, but we will participate via an aggregator’

One participant reasoned that direct participation in the TA was no longer worthwhile but they would participate in the first TA via an aggregator. This participant indicated in interview that they did not feel it was worth making a major investment to upgrade metering equipment so that their existing generation CMU could meet metering test requirements for all its capacity. They chose to withdraw this CMU and submit capacity via an aggregator’s unproven DSR CMU (which allowed them to select cost-effective sites).

Reasoning: ‘The TA is more demanding than we expected and it’s no longer worthwhile to participate’

Another participant withdrew a DSR CMU that did not meet the 2 MW threshold for participation. Their decision may have been influenced by their view that baseline requirements[15] for DSR were problematic for assets that were also participating in frequency services. This contrasted with the views of other participants who reported being able to combine delivery for the TA and frequency services: we are not yet clear what explains these differing views.

[15] Baseline requirements for DSR are explained in Appendix 1. They involve readings measured over a six week period prior to the test or stress event, in contrast to more limited STOR baseline requirements.

3 Fulfilment of obligations in first TA

Fulfilment requirements

Participants with CMUs in the TA must deliver against their capacity obligation at any time of system stress during the Delivery Year, or face a financial penalty. CMNs are issued by National Grid when a shortage of generation is anticipated. Capacity providers do not receive despatch instructions. Instead, the CMN is a signal to all providers that system stress is anticipated (although it may not materialise). If a stress event occurs, four hours after the issue of the CMN, any participant who fails to deliver their capacity obligation is subject to penalties.

At the time of this research, two CMNs had been issued but neither led to a stress event:

- CMN1 was issued on 31st October at 12.06pm, and was live from 4.30pm to 7.00pm. The cancellation notice was issued at 6.53pm.
- CMN2 was issued on 7th November at 12.06pm. This was due to go live at 4.30pm but was cancelled at 3.07pm.

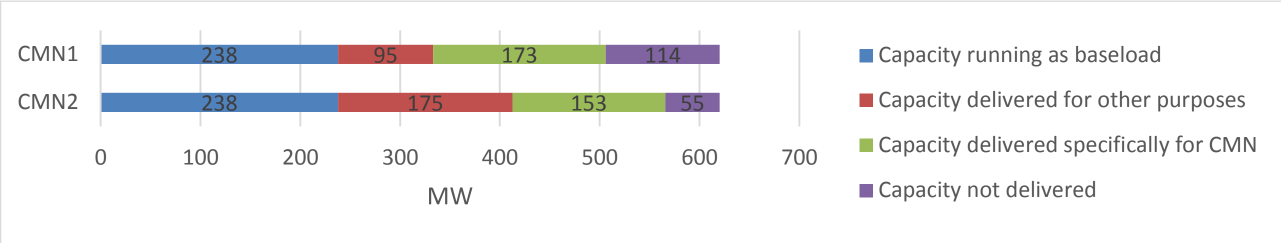
We proposed an initial theory as to how TA participants would fulfil their obligations in response to a stress event, as shown in Appendix 2. We have revised the theory based on interview evidence, including participants' reported responses to the two CMNs. Our revised theory, shown in Appendix 3, focuses on CMN responses as no stress event has yet been observed.

Summary of fulfilment outcomes for first TA

National Grid holds no meter data about the CMN responses because no stress event occurred. Our analysis of CMN responses is therefore based on interview data from Phase 2, supplemented in a few cases by information on CMU characteristics from Phase 1.

Our estimate of the proportion of capacity responding to the CMNs, shown in Figure 3.1, is based on the instructions that aggregators gave their clients. Evidence from aggregator client interviews suggests that responses were generally consistent with the aggregator's instructions, although clients could choose not to respond for operational reasons. In a few cases, clients prepared to respond but changed their plans when their aggregator advised that the CMN had been cancelled. More capacity was already being delivered for other purposes for the second CMN, largely because this CMN coincided with Triad warnings. The reasons for different reactions are explained below.

Figure 3.1 Estimated delivery in response to CMNs, as reported by participants in interview



How, why and for whom these outcomes arose

All interviewees reported that they had been aware that the CMNs had been issued and had some recall of their response, although some had limited recall. Subject to this caveat, the reasonings underlying the outcomes in Figure 3.1 are explained below.

Reasoning: ‘No change to normal operation as we are running at full capacity anyway’

Seven participants who offered small-scale generation that was running as baseload (nearly 240 MW, across 13 CMUs), were already providing their full capacity, or close to their full capacity, at the time of both CMNs. This was nearly 40% of TA capacity.

This was because, barring technical problems or maintenance closures, baseload generation was always running. Some of these participants commented that they planned maintenance periods for the summer months, to reduce the risk of maintenance closures coinciding with system stress. One commented that, at the time of the CMNs, they made sure they were delivering sufficient capacity to meet their obligation.

Reasoning: ‘No change required as already responding for other purposes’

A number of participants with CMUs offering standby generation or turn-down DSR were already providing their full capacity for other purposes at the time of the CMNs. We estimate that this applied to six participants (13 CMUs, 95 MW) for the first CMN and seven participants (16 CMUs, 175 MW) for the second, representing 15-30% of TA capacity. They did not change their delivery in response to each CMN, but monitored the situation in case they needed to sustain delivery solely for the CMN.

These participants had non-baseload assets and were confidently active in the flexibility market. Specific contexts in which this reasoning applied included: peaking generators exporting to the grid because of high electricity prices; participants responding to Triad warnings (particularly for the second CMN, which was within the Triad period); participants ‘held in readiness’ for other flexibility services such as STOR or FFR (which counted as full delivery); and participants planning to use the CMN period as one of their SPDs.

Reasoning: ‘Respond to CMN because risk averse/think a stress event might occur’
Some aggregators who offered standby generation or turn-down DSR, and were not already delivering for other services, chose to deliver ‘just in case’ a stress event occurred. We estimate that there were 5 such participants (14 CMUs, 173 MW) for the first CMN, and 4 such participants (12 CMUs, 153 MW) for the second, representing 25-30% of capacity.

The main context here was that these participants were not fully confident in their understanding of the electricity system. Some participants in this group commented that their current strategy is to respond to all CMNs, but their strategy might be revised in future, if it was not cost-effective because of the number and length of CMNs.

> *At present that [responding to the CMN] is the safest option. [The] precautionary principle applies. (Aggregator)*

Reasoning: 'Ready to respond but do not take any action because not needed yet’
A few aggregators and direct participants did not take any action in response to the CMNs and chose to monitor the situation. We estimate that they comprised four organisations for the first CMN (6 CMUs, 114 MW) and a (different) four for the second (5 CMUs, 53 MW). This represented 18% of TA capacity for the first CMN and 9% for the second.

The key context here was that these participants were confident that they understood the electricity system and could respond quickly if needed. They were not delivering baseload and, at the time of the CMNs, happened not to be delivering for other flexibility services or for Triad. Their view, from monitoring the electricity system, was that these two CMNs were unlikely to result in stress events. One provider commented that they judged that the second CMN was less likely to develop into a stress event than the first CMN. Two DSR providers commented that they chose a ‘wait and see’ policy because they could respond quickly and an early response might adversely affect their baseline for a later stress event.

> *“[We] did not respond - It’s only a warning, and it’s just like, “Alert. You need to be cognisant of this and you need to monitor things.” It doesn’t mean you need to start turning down things, because if anything, if you start turning down things immediately, you get caught out by the baseline.” (Aggregator)*

4 Contribution to security of supply in 2016/17

We tested whether the TA contributed to its objective (H1) of contributing to security of supply between October 2016 and September 2017. This was assessed in terms of whether capacity was made newly available (or kept available) to National Grid, that would not have been available without the TA[16]. This assessment was unaffected by National Grid's usage of or dependence on this capacity in practice.

Hypothesis 1: The TA leads to direct participants and aggregators making or keeping capacity available that would otherwise have been closed/mothballed or that was not part of a formal network balancing service. This capacity contributes to security of supply and/or meeting the reliability standard in the 2016/17 delivery window.[17]

We developed an initial theory about TA participants' contribution to H1, as shown in Appendix 2. This theory remains broadly valid in the light of findings below.

Summary of security of supply outcomes for first TA

Our analysis of CMU data, auction data and interview data suggests that an estimated 20-30% of the capacity in the first TA would not have been made available to National Grid in 2016/17 without the TA. Between 130 MW and 180 MW of capacity across 23 CMUs was 'additional' in this sense, comprising about two-thirds DSR and one-third existing generation. This estimate is consistent with the preliminary estimate of 29% in Phase 1.

These estimates are based on interview statements on the proportion of each CMU's capacity that depended on TA revenues. Broad ranges are given here because of limited data about the proportion of 'additional' capacity in some CMUs.

This estimate is slightly lower than the estimated 44% of CMUs providing likely support for H1 in contribution tracing analysis (see Appendix 5), because contribution tracing identifies

[16] Previous participation in Triad targeting and Demand Side Balancing Response (DSBR) were not counted as current National Grid services, because Triad is not controlled by National Grid and DSBR was not tendered in 2016/17.
[17] The actual use of this capacity to deliver services to the TA, to Triad management and/or other balancing services in 2016/17 does not affect the assessment of additionality: the hypothesis is simply that additional capacity is made (or kept) available that would not have been available without the TA.

whole CMUs as ‘additional’ or ‘not additional’, whereas interview evidence enabled us to estimate the proportion of each CMU that is ‘additional’.

In all cases, interview evidence indicated that TA revenues were being stacked with other sources of revenue (e.g. Triad, STOR, FFR and/or sales to the electricity market), both in the current year and in future years. Interview evidence indicated that aggregator and client decisions to invest in new capacity, controls or equipment were premised on expected revenue streams over several years, not just one year of TA revenues.

How, why and for whom these outcomes arose

The main reasoning for these outcomes is presented first.

Reasoning: ‘The TA is a no-brainer as we will be providing capacity anyway’
Most (70-80%) of TA capacity was already available to National Grid flexibility services. Between 440 MW and 490 MW of capacity across 41 CMUs[18] was ‘non-additional’ in this sense, comprising about half existing generation and half DSR. The contexts were:

- Existing generation (provided by baseload or peaking/back-up[19] generation, via an aggregator or direct) that was already contributing to a National Grid service and that could stack TA revenue with other compatible services (e.g. STOR, FFR).

- Unproven DSR capacity (back-up or turn-down assets, via an aggregator or direct participant) that was already contributing to a National Grid service and could reliably demonstrate capacity relative to its baseline while stacking TA revenue with at least one other compatible service (e.g. STOR, FFR, Triad).

Reasoning: ‘While we would provide some capacity without the TA, we have expanded the capacity we provide, primarily because of the additional financial incentive of the TA ‘
Six participants expanded the capacity they provided because of additional TA funding, where they could comply with the conditions of CM participation. The contexts were that:

- Five were DSR aggregators with an existing client base in another flexibility service which was compatible with and could be stacked with the TA, with the ambition and capability to encourage their existing clients to bring forward additional sites and to

[18] This includes some CMUs which were partially additional, for a proportion of their capacity, because some sites or some clients in the CMU were new to flexibility.
[19] The primary role of back-up generation was providing onsite resilience, while the primary role of peaking generation was generating for flexibility services. Some back-up generation was classified as existing generation, and some as DSR.

recruit clients new to flexibility services. One client commented that the longer lead time for CMNs allowed the addition of sites that could not deliver for other services.

> *"Yes, there is [capacity volume] in the Capacity Market that isn't in STOR from our asset base. [...] We couldn't participate in STOR with that [asset], but we can participate in the Capacity Market because four hours [notice] is sufficient. (Aggregator client)*

- One was a direct participant with multiple sites where the TA allowed them to bring more sites into the flexibility market by funding initial investment in controls or grid connections for these sites.

Factors that were identified by interviewees as influencing the scope for making additional capacity available for the 2016/17 delivery year included:

- Whether capacity was classified as existing generation or unproven DSR. Back-up generation used both classifications, but only the unproven DSR classification allowed flexibility in the amount of capacity proved for the 2016/17 delivery year, with low penalties for reduced volumes in the TA in terms of lost credit cover.
- The time required to bring new clients on board and meet testing requirements, compared to the short time-scale between the first TA auction and start of the delivery year, which was particularly challenging for clients new to flexibility.
- The level of investment required to bring an additional site into the market (depending on the investment required in controls, grid connection and synchronisation equipment[20]) relative to the scale of capacity available on that site. Some participants reported that at least 100-300 kW was typically required to justify the enabling investment and hassle of bringing a new site into the flexibility market.

Aggregators differed in their success rates in bringing on board new clients and new sites for delivery in 2016/17. Successful participants appear to have made a significant investment in marketing to new clients, and also to have developed an early understanding that only sites with supplier settlement metering and no renewable generation could feasibly be signed up in time for delivery in 2016/17.

[20] Costs were high for sites offering back-up generation where these required a new DNO connection and synchronisation equipment. One client mentioned that it could cost £10,000 – £50,000 to bring a new back-up generation site into the flexibility market, premised on future revenues from TA/CM and other services.

Reasoning: 'Without the TA we would not be providing capacity to National Grid, but TA funding, stacked with other revenue, gives us the financial incentive to do so'

The TA prompted five participants to provide 'additional' capacity to National Grid where:

- Two were direct participants with insufficient revenue from other sources to maintain capacity, from generation assets that were too uncertain to participate in longer horizon services such as T-4.
- Two were direct participants with generation or DSR new to the flexibility market that were using the TA as a testing ground for the CM and other flexibility services.

"We haven't participated in any other balancing market schemes, such as STOR, or FFR, etc., our site is currently just operating on base load power price, plus the Capacity Market." (Direct participant)

- Another was an aggregator new to the market in the GB, attracted by TA/CM opportunities and seeking to develop new DSR portfolios, who saw the TA as the initial offer for new clients, to which they would add other services in future.

…the TA is essential because there's an opportunity for an aggregator to pursue, with some money attached, that allows us to go out and talk to people." (Aggregator)

5 Contribution to development of capacity for main CM

We tested whether the TA contributed to its objective (H2) of leading to more, or more competitive, capacity for the CM in 2017/18 and future years. We assessed this in terms of whether the TA had encouraged the development of capacity for the main CM (e.g. the T-4, T-1 or Early Auction) rather than the second TA, as the latter will not be available beyond 2017/18.

Hypothesis 2: The first TA leads to more (competitive) capacity for the CM in 2017/18 and subsequent years.

We developed an initial theory about the TA participant's contribution to H2, as shown in Appendix 2. This theory remains broadly valid in light of the findings below.

Summary of TA outcomes for future CM capacity

There was consistent evidence that the TA had acted as a 'pilot' for the main CM, not just for TA participants but also for National Grid, EMRS and for BEIS. As intended by BEIS, it generated significant learning about how CM testing and delivery work in practice.

Ten participants reported in interview that their TA experience had encouraged them to provide more (or more competitive) capacity in the main Capacity Market for 2017/18 or beyond, than they would have done without the TA. This included some organisations that dropped out from delivery in 2016/17 but had learnt from their TA experiences.

Seven participants felt better prepared for the main CM because of their experience in the TA but reported that the scale and competitiveness of their future capacity was not significantly influenced by the TA.

Seven participants gained a better understanding of CM rules through the TA, but reported that this had made them more cautious about providing capacity in the main CM.

These findings were broadly consistent with contribution tracing findings presented in Appendix 5, which found that 18 out of 23 participants[21] were 'likely' or 'very likely' to

[21] As detailed in Appendix 5, all 24 participants in the first TA were subject to a 'screening test' before the contribution tracing analysis was conducted, of which 23 participants passed

support the hypothesis that the first TA contributed to the development of capacity for the main CM. But 12 out of 23 participants also supported the competing hypothesis that participants would have bid into the future CM auctions at similar prices, irrespective of the first TA. The cases supporting both hypotheses were those where the TA generated learning but did not significantly affect the competitiveness of the capacity they expect to offer to the main CM in future.

These findings were slightly less positive than Phase 1 findings on H2, because many participants had found testing requirements challenging, and some were concerned about the stability of future CM revenues given the low clearing price in the Early Auction.

How, why and for whom these outcomes arose

Positive outcomes for this hypothesis were dependent on participants' confidence in the CM as a future revenue stream, at least at the time of TA investment, in combination with other sources of revenue. The first four reasonings below were associated with positive outcomes for this hypothesis, while the remaining three led to negative outcomes.

Reasoning: 'We invested in capacity for the TA which will make us better-positioned to participate in the CM in the future.'
Three participants reported that they had invested in physical assets, in their client base or in strategic learning which would allow them to participate more competitively in future CM auctions. Contexts for this reasoning were:

- Established aggregators with an existing client base who marketed actively to expand their client base by offering TA revenue, stacked with other opportunities.
- Aggregators, direct participants and aggregator clients with sites that did not meet CM metering accuracy requirements, who invested in metering or transformer equipment to meet TA requirements, which they could also use in the main CM.
- Aggregators who overcame unanticipated costs and challenges in getting their CMUs proven for the TA, and adapted their strategies to make future CM participation more cost-effective (e.g. focusing on sites with supplier settlement metering and no renewables).

Reasoning: 'We wanted to enter the CM and the TA provided a low risk way to build our customer base'
A further four participants who were new entrants to the UK aggregation market used the TA to build their customer base for the main CM. Relevant contexts were:

- For DSR aggregators with experience in other countries and an interest in entering the main CM, the TA was a low-risk stepping stone to the wider CM, providing a base income to attract clients, on which to layer other flexibility services.

- For energy suppliers in GB who were successful in developing their flexibility offer for their existing electricity supply customers, the TA was a low-risk stepping stone towards offering flexibility services, including the CM.

Reasoning: 'Now we have experience of the TA, the CM seems less risky.'
In addition, three TA participants reported in interview that they had gained confidence in the CM through their experience of the TA. Contexts associated with this reasoning were:

- Established aggregators with existing clients offering generation and/or mixed DSR who saw the CM as a strategic opportunity but were not fully familiar with CM rules and their implications.
- Direct participants attracted by the TA's low risk environment (e.g. low credit cover) who might not otherwise have risked direct participation in the main CM.

> *There's numerous opportunities out there, which we actively continue to appraise, and the knowledge we've gained through participation in the TA and the Early Auction, give us a much greater understanding of that process, and how the revenue streams work. (Direct participant)*

Reasoning: 'The TA has been a useful test run for the main CM, but we were always going to participate in the CM.'
Seven participants thought that the TA had generated useful learning but had not significantly influenced the capacity that they would offer the main CM in future.

These participants were confident about the CM and held positive views about future CM opportunities. Many were established aggregators putting forward generation CMUs and/or mixed DSR CMUs (i.e. back-up and turn-down) from their existing clients.

Reasoning: 'The TA was harder than we thought – we won't participate in the CM again.'
Two participants had problematic TA experiences which had put them off future CM participation, although they would continue in other flexibility services.

One was an aggregator who was not successful at signing up new clients as required to fill their unproven DSR CMUs (e.g. because of the short timescale for sign-up and some critical processes being outside their control – e.g. dependence on energy suppliers setting up dataflows or providing metering information).

> *We've had a few painful lessons. Compared to the expectations that we have, I cannot really describe it as a success and so that is a significant*

issue. I would say that… we are more cautious about putting anything to the Capacity Market… given the experience that we've had. (Aggregator)

Another potential aggregator reported that their TA experience had led them to focus on putting generation rather than DSR into the CM in future, because it had influenced their perception about the size and profitability about the DSR market.

The company view is that DSR might not … be going to provide us with a sustainable model. We've put more money into it than we will ever get out of it in the next couple of years and you would need a huge amount to sign up [to be profitable]. (Aggregator)

Reasoning: 'The TA was harder than we thought – we will only participate as aggregator clients in future.'

Two participants found the TA hard but planned to participate via an aggregator in future.

One of these was a direct participant who experienced problems meeting metering requirements for some sites, partly because of renewable generation on these sites. As indicated in section 2, they chose to withdraw and participate via an aggregator instead. They reported that the TA had made them more aware of the challenges of direct participation in CM.

We have indirect evidence that at least one of the potential aggregators who has a client base for other energy services decided to collaborate with another more experienced aggregator for their clients' capacity, rather than acting as an aggregator themselves. But we have no direct evidence about the contexts influencing this choice.

Reasoning: 'The conditions of the main CM are less attractive to us than the TA.'

Three direct participants reported that they were unlikely to participate in the main CM, because the conditions were less attractive, despite successful participation in the TA. The aspects of the main CM that were unattractive to them included higher credit cover for unproven capacity, termination fees and the expectation of lower prices (particularly given the low clearing price in the Early Auction). Contexts that influenced participants' reasoning in some of these cases were that:

- All three offered turn-down DSR or generation with significant running costs (so they might look for a higher clearing price to justify providing capacity).
- Some had assets suitable for participation in frequency services, and perceived these services to conflict with CM delivery (e.g. because of DSR baseline issues).

- Based on their TA experience, some were concerned about the complexity of CM rules and the risk of non-compliance (e.g. because of uncertainties in interpretation of the rules).

6 Contribution to encouragement of turn-down DSR

We tested whether the TA contributed to its objective (H3) of encouraging turn-down DSR, within and beyond the CM. We assessed this contribution primarily in terms of the first TA, but are aware that the second TA is focused solely on turn-down DSR and will make a greater contribution to this objective.

Hypothesis 3: The first TA leads to wider encouragement of turn-down DSR

We developed an initial theory about TA participants' contribution to H3, as shown in Appendix 2. The revised theory in Appendix 3 takes account of the findings below.

Summary of TA outcomes for turn-down DSR

Interview evidence, combined with data on CMU characteristics, indicates that only 10-15% of capacity delivered for the first TA was turn-down DSR. This was slightly lower than the estimate of 19% for Phase 1 because final sign-up of turn-down DSR clients by aggregators was lower than expected at the time of Phase 1 research. This capacity comprised just under 30 MW of turn-down DSR from direct participants, and a further 30-60 MW[22] of turn-down DSR across 16 DSR CMUs put forward by aggregators. Evidence from interviews with aggregator clients indicates that some organisations were encouraged by the TA to offer turn-down DSR for the first time.

These findings were broadly consistent with contribution tracing findings. Of the 16 participants reporting some interest in turn-down DSR, only five cases showed evidence that they were 'very likely' or 'likely' to support H3 (i.e. that the TA provided wider encouragement for turn-down DSR). But only three of these 16 cases were 'very likely' or 'likely' to support the competing hypothesis that 'turn-down DSR was seen as a long-term opportunity, even if not currently cost-effective, and TA was not needed to encourage this'. The implication, as explained below, was that many participants felt that turn-down DSR required more support than that provided by the first TA.

[22] This is a rough estimate based on interview statements, not direct data.

How, why and for whom these outcomes arose

There was unanimous agreement amongst aggregators that it was more difficult to recruit clients offering turn-down than back-up DSR, because turn-down was perceived as potentially conflicting with an organisation's main business activity.

> *My personal view is that [electricity] generation and storage is always going to be the route preferred by the majority of […] DSR providers, because, whilst there are some issues around getting connections sorted out and stuff like that, it has a much lower level of interference with the rest of the operation of your business. (Aggregator)*
>
> *"Whilst we do want to participate in these markets, and we are looking at additional revenue streams, we can't do it whilst compromising our core business." (Aggregator client)*

The process of getting a turn-down client on board was reported to be more time-consuming than for back-up generation, because senior management needed to assess the risks to their business. This made recruitment particularly challenging given the tight timeframe between the first TA auction (when the clearing price was known) and the start of the delivery year.

However, a few participants also noted that prospects for DSR from diesel generation, as an alternative to turn-down, may be adversely affected in future by the Medium Combustion Plant Directive.

The reasonings below start with organisations most committed to turn-down DSR.

Reasoning: 'We've always done as much turn-down DSR as we can'
Two organisations, one aggregator and one direct participant, reported that they were strongly committed to turn-down DSR but already did as much as they could. Contexts for this reasoning were:

- The direct participant had loads highly suitable for turn-down (see below) but were already maximising their turn-down DSR opportunities via Triad and flexibility services, within the operational constraints of their business.

- The aggregator was committed to, and skilled at, turn-down DSR and did not see the CM as an attractive route for their turn-down assets compared to other services (e.g. frequency response). There is some evidence (e.g. from marketing materials) that this may be explained by the type of assets offered (e.g. HVAC or refrigeration processes offering automatic turn-down for short periods, rather than assets able to turn-down for longer periods). This was consistent with Phase 1 findings that some non-TA aggregators saw a conflict between providing evidence of a baseline for turn-down DSR and delivering frequency services.

> *We do not currently participate in the CM with turn-down DSR. [...] we believe the baselining, testing and evidence of delivery for DSR is ambiguous." (Aggregator)*

Reasoning: 'The TA has prompted us to offer turn-down DSR for the first time.'
Some aggregator clients had been prompted by a deal with an aggregator to offer turn-down DSR for the first time to the TA, usually as part of a multi-year deal including the first and second TA. We found direct evidence of two such clients, and indirect evidence that aggregators were recruiting some new turn-down clients (see below).

> *"It [the TA] has prompted us to do something [turn-down DSR] that we wouldn't have otherwise done, so it is something that we wouldn't have otherwise been offering." (Aggregator client)*

Contexts for this reasoning were that:

- They had loads suitable for turn-down for periods longer than 30 minutes (e.g. equipment involved in industrial, production or storage processes).

> *"We pretty much know what customers can switch off their processes, we know like water treatment, water distribution; we know that cold storage, steel manufacturing, cement [can do it] as well." (Aggregator)*

- The scale of load reduction was sufficient to justify any investment required in metering or controls (e.g. 100 kW or more per site, depending metering)[23].
- They expected a low number of turn-down requests[24] and were confident that they would not need to turn down in a way that put their main business at risk (e.g. they could choose not to turn down in response to particular CMN).

[23] Grid connections and synchronisation equipment were not required for turn-down assets. Costs were low where sites were able to use supplier settlement metering, but could be several thousand pounds where other metering options had to be used (e.g. because of onsite renewables).

- They did not feel they had the knowledge and expertise required to participate directly in the CM, even if their load exceeded the minimum threshold of 2 MW, so chose to participate via an aggregator. For example, some previously targeted Triad but had not been involved in other flexibility services.

- Some, but not all, had Corporate Social Responsibility (CSR) or carbon reduction objectives that supported turn-down.

Reasoning: 'We were already committed to turn-down DSR, but the (first/second) TA has enabled us to do more'

Five aggregators and direct participants were already strongly committed to providing turn-down DSR, but the financial incentive offered by the first TA, combined with the prospect of taking turn-down assets into the second TA, provided an additional incentive for them to find more turn-down DSR capacity. In addition to the client contexts outlined above, additional contexts for this reasoning were:

- Aggregator knowledge and skills in turn-down DSR (e.g. software platform).

- Aggregator or direct participant commitment to turn-down DSR (driven by CSR or carbon reduction objectives or by the suitability of their assets and systems).

- Aggregators with clients offering back-up generation, who could be approached to offer turn-down as well.

Reasoning: 'The TA has encouraged us to enter DSR aggregation, including turn-down aggregation'

Three organisations were encouraged by the TA to enter the DSR aggregation business in the UK, and provided an element of new turn-down DSR within their portfolios. These aggregators were experienced in turn-down DSR in other countries but new to the GB aggregation market. They saw aggregation of DSR in the main CM as a long-term opportunity and the TA as a low-risk entry route. These aggregators offered new turn-down aggregation as one part of the new capacity they signed up for the TA.

These new aggregators looked for turn-down opportunities from new clients, choosing different marketing strategies depending on their wider offer (e.g. software platforms): some targeted large-scale loads in particular sectors while others approached smaller organisations new to flexibility which had at least some potential for turn-down.

[24] There was considerable variation in the reported expectations of TA participants, from 20-30 hours of stress events per year (similar to Triad and STOR) down to 2-3 hours of stress events per year. It is not clear how these expectations relate to their level of understanding of the electricity system.

> *...the big untapped potential for having lots of turn down [DSR], is from organisations that aren't that big and don't have full time energy people. (Aggregator)*

Reasoning: 'We would like to do more turn-down DSR but it's challenging and needs more support than offered by the (first) TA'

Five participants reported that the first TA did not provide an adequate timeframe or incentive to encourage significant volumes of turn-down DSR, in spite of their interest in providing more turn-down DSR. In some cases, these aggregators still planned to offer turn-down DSR portfolios for the second TA, as this offered a longer time-frame. Contexts influencing this reasoning were:

- Aggregators who had less experience of turn-down DSR, had no particular commitment to turn-down, and who found it difficult to recruit turn-down DSR for the first TA (e.g. owing to perceived operational issues for clients' core business).
- Aggregators who were experienced and committed to turn-down, but who faced challenges in signing up new turn-down DSR capacity within the tight timeframe for the first TA and saw back-up generation as the 'low hanging fruit' for the first TA.
- Direct participants and potential clients with assets potentially suitable for turn-down who were concerned about the potentially unlimited number and length of stress events in the TA/CM, which could make the economics of turn-down marginal.
- The perceptions of some aggregators that there was a relatively limited pool of clients with assets suitable for longer turn-down periods.

Reasoning: 'We participated in the TA on non-turn-down DSR projects, and we have no interest in or capacity for turn-down DSR'

For eight TA participants, the first TA had no impact on their interest in or capacity for turn-down DSR because the organisation's business model was entirely based around generation. Contexts for this reasoning included:

- Aggregators involved in the aggregation of peaking generation assets, who did not have knowledge or skills in turn-down DSR aggregation.
- Direct participants offering peaking generation or back-up generation, who did not have loads suitable for turn-down DSR, or who regarded turn-down DSR as posing unacceptable risks for their business.

"People who have the flexibility, generation sets or processes which are quite easy to switch off, are the easiest ones, and we just need to identify them. But some others, they are not particularly interested in doing it, because their switching of the process may be too costly, or there's too much risk associated with it." (Aggregator)

7 Value-for-money of the first TA

Phase 1 findings on value-for-money

A full value-for-money assessment of the first TA is beyond the scope of this evaluation and would need to compare the cost of the TA with the cost of achieving TA objectives by other means (e.g. other means to bring forward and encourage DSR).

Nevertheless, during Phase 1 of the evaluation, we made a preliminary assessment of value-for-money, including an assessment of the first TA's contribution to its objectives. We modelled costs for different types of TA capacity, using industry data[25] supplemented by cost information gathered through interviews, including estimates of capital and operating costs directly attributable to TA participation. We also modelled the revenues that TA capacity would receive from other sources, including electricity sales and flexibility services, using interview data about running hours and participation in flexibility services. Using these costs and revenues, we modelled the 'missing money' that participants with different CMUs would be expected to seek to recover from the TA. In cases where interviewees made clear that capacity was already available to the market, and the TA was simply topping up existing revenues, we assumed they would only seek to recover costs specific to TA participation. The estimates of 'missing money' were used to compile a supply curve for the first TA auction, which was compared to the auction demand curve to generate a theoretical 'supply curve clearing price'.

Our findings from Phase 1 were that the first TA auction had made some contribution to security of supply in 2016/17, and appeared likely to bring forward some capacity for future CM auctions. However, the clearing price of £27.50/kW was high relative to our estimates of underlying supply costs[26]. This is relevant when considering the cost effectiveness of the auction.

[25] We modelled capital expenditure, operating costs, maintenance costs, fuel costs and carbon costs. Diesel generation costs were sourced from: Leigh Fisher Jacobs (2015). *Electricity generation costs and hurdle rates. Lot 3: non-renewable technologies.* Prepared for DECC, February 2015. Draft.

[26] The supply costs that were included in the earlier analysis, and the limitations of the Phase 1 value-for-money analysis, are detailed in section 3 of the Phase 1 report: https://www.gov.uk/government/publications/evaluation-of-the-transitional-arrangements-phase-1

Updated information from Phase 2

New cost evidence from Phase 2 is broadly consistent with the supply curve modelling undertaken during Phase 1. Additional information was gathered on:

- The characterisation of the CMUs going forward to delivery for the first TA (e.g. proportion of turn-down; fuel source; final volume of CMU).
- The running hours involved in responding to DSR tests, SPDs and CMNs during the current delivery year, when capacity would not be delivered for other purposes.
- Cost incurred in signing up clients and meeting testing requirements (including staff time and capital costs for new metering or equipment).

We have gathered little additional data on the opportunity costs of turn-down DSR during Phase 2, as these are both confidential and highly variable. These costs depend on the nature of the industry offering turn-down and on the timing of a specific CMN relative to an organisation's normal activities. We will consider how to address this in Phase 3 of the evaluation.

Implications for value-for-money

In our Phase 1 assessment of value-for-money, we compared the first TA's contribution to its objectives with the costs associated with TA implementation. As explained above, the new cost data collected during Phase 2 is broadly consistent with the supply curve modelling undertaken during Phase 1.

Similarly, there has been little change in our assessment of the first TA's contribution to its objectives, which forms part of our overall assessment of value-for-money although Phase 2 findings, as detailed in sections 4, 5 and 6 of this report, suggest that this contribution was marginally more limited than previously estimated in Phase 1:

- For many participants, the first TA still contributed significantly to its objective of encouraging more DSR and small-scale generation to come forward for future CM auctions, but, as evidenced in section 5 of this report, a few participants had more negative views about participating with DSR in the main CM as a result of problematic TA experiences.
- Our revised estimate of the first TA's contribution to encouraging turn-down DSR was also slightly lower than estimated in Phase 1, as evidenced in section 6 of this report. This was primarily because of the challenge that aggregators faced in recruiting new turn-down clients within the time-scales of the first TA.

- Our assessment of the TA's contribution to security of supply in 2016/17, as evidenced in section 4 of this report, was largely unchanged compared to Phase 1.

Overall, these Phase 2 findings do not significantly change our Phase 1 assessment of value-for-money of the first TA auction.

8 Wider learning for the CM and DSR

This section highlights learning points which have implications beyond the TA scheme itself, and explore the implications of evaluation findings for the wider CM and for future policy on DSR and aggregation.

Consistency of CM and other energy objectives

There was considerable evidence, from many participants and delivery partners, that CM metering accuracy requirements made it difficult for complex sites with renewable energy generation to participate in the CM. The metering accuracy required by the CM is more demanding than the accuracy required for other flexibility services or for Feed-in-Tariff or Renewable Heat Incentive projects. This was a source of frustration for industry and acted as a barrier to participation of DSR in the CM.

> *It is surprising that for electricity generation as a whole there isn't one set of electricity metering requirements that makes them [assets] comply for everything. Very surprising. (Aggregator client)*

There was also some suggestion, less well evidenced, that participants experienced conflicts between energy efficiency and flexibility objectives. For example, an energy manager with a limited budget of time and money, at least in the short-term, may effectively have to choose whether to invest their time/budget in a flexibility or energy efficiency initiative.

The reliability of DSR

Our Phase 1 report found that the low credit cover for the first TA encouraged aggregators to take the risk of submitting large, unfilled CMUs of unproven DSR. In Phase 2, we observed significant reductions in capacity post-auction, as some less experienced aggregators failed to fill their CMUs. Nevertheless, the low risk environment provided by the TA encouraged new entrants and allowed participants to test new approaches to DSR.

It was too early to assess the reliability of DSR for the electricity grid, or its potential contribution to National Grid’s security of supply, as there have not yet been any stress events. There were wide variations in the number, size and type of components in each CMU, from 1 to around 50. Case study analysis of meter data for DSR tests and SPDs suggested that the contribution of components varied between tests, particularly for aggregator CMUs with large numbers of components, but that most of the capacity was delivered by a small number of larger components. The implications of portfolio design for DSR reliability, for example whether having a portfolio with a large number of small capacity components improves reliability, will be explored in later phases of the evaluation

As discussed in section 3 above, there was some suggestion that the DSR baseline methodology acted as a disincentive for DSR providers to deliver capacity early after a CMN, if they expected it to develop into a stress event. Depending on the accuracy of a participant's understanding of the electricity system, this might increase the risk of late delivery during a stress event.

Typology for turn-down DSR

Phase 2 evidence suggested that some types of turn-down DSR could generate attractive returns through frequency response services. Some participants perceived that this may not be fully compatible with delivery for the CM because of baselining issues. Further research is required in Phase 3 to establish whether the conflict between frequency response and CM participation is perceived or real, particularly since there have recently been changes to the CM rules about delivery for frequency response.

Phase 2 evidence suggested that some DSR assets were better suited to the CM rather than frequency response and vice versa. Figure 8.1 characterises the types of DSR assets that appeared suited to these two flexibility services.

While it was theoretically possible for aggregators to cover long stress events by coordinating delivery from multiple assets that can only tolerate short periods of turn-down, we did not find evidence of this happening in practice. One aggregator commented that 'overfilling' a CMU was commercially unattractive because each client would receive a lower share of TA revenues.

Table 8.1 Preliminary typology of turn-down DSR

Characteristics	Suitable for CM	Suitable for frequency response
Examples of typical processes	Industrial processes with spare capacity or potential for storage; water pumping; cold stores	HVAC, refrigeration, heating, 24/7 production of perishable products.
Significant notice period required	✓	×
Fast response	✓	✓
User control	✓	×
Automatic response	✓	✓
Capable of long turn down periods	✓	✓
Only tolerate short turn-down periods (e.g. less than 30 minutes)	× (unless aggregated with other short turn-down loads)	✓

Source: TA evaluation evidence.

Aggregation vs direct participation

Aggregators played a major role in the first TA, using TA incentives to bring new clients into the flexibility market. While Phase 1 found that the TA initially encouraged the entry of six new players into DSR aggregation for the CM in the UK, we found in Phase 2 that two have reviewed their future strategy and report that they are withdrawing from this market. The Phase 2 findings are corroborated by press reports suggesting that some form of consolidation is likely[27]. Linkages were reported to be developing between energy suppliers and aggregators: energy suppliers offer an existing client base with ready access to meter data, while aggregators offer software platforms and specialist knowledge of the DSR market. At the time of writing, the Association for Decentralised Energy was

[27] http://theenergyst.com/demand-side-response-aggregator-market-faces-significant-consolidation/

consulting on a draft code of practice that would set standards for aggregators' interactions with clients and would also affect this market.[28]

While direct participants have also played an important role in the first TA, they have mainly participated in generation rather than turn-down DSR. Phase 3 of the evaluation will explore in more detail why the second TA has not seen growth in direct participation by turn-down DSR (e.g. complexity of CM rules; perceived risks to main business of providing turn-down; lack of confidence about the flexibility market and DSR; and/or alternative attraction of frequency services for certain kinds of assets). We have developed theory on the contexts influencing organisations to participate via an aggregator, as presented in Appendix 3.

CM rules and their implications

Timescales

The short timescale between CM auctions and the start of the delivery year affected sign-up rates for the TA. This is also likely to be an issue for the second TA, which has a tighter timetable, and could act as a barrier to entry for future T-1 auctions. This is consistent with observed participation by DSR in recent T-4 auctions, which offer a longer lead-time for aggregators that have access to other revenue streams in the intervening years.

Complexity and need for better guidance

In addition to the accuracy point raised above, many TA participants commented that CM rules on metering and DSR testing were complex, unclear and challenging, with long and inadequate guidance and uncertain interpretation.

> *It would be nice to have a precise version; "if you have this type of asset, you need to look at x, y, z." … There's so much documentation that I never got to read because there is too much of it. [I] got to 2,000 pages and gave up counting. (Direct participant)*

Many felt that the complexity of the rules, and the uncertainty of their interpretation, were barriers to entry into the CM and lent an advantage to the more experienced players. At the time of writing, Ofgem and National Grid were working to address these issues.

[28] https://www.theade.co.uk/assets/docs/nws/DSR_CoC_Consultation_Document_-_Final_-_18_July_2017.pdf?mc_cid=faa03e0f83&mc_eid=2f23ad2b98

Process issues

Interviews and case studies highlighted the views of many participants that the flow of communication between the different bodies – National Grid, EMRS and the Meter Operator (MOP) – was overly convoluted. Several complained that National Grid did not meet their regulatory timescales, causing issues for those that were eager to hear their test results and have time to organise retests. One participant suggested that there should be an account manager at National Grid for each participant.

Consideration could also be given as to how to improve the engagement and cooperation of DNOs and suppliers, to overcome the issues with reliance on external parties identified in this report.

Comments on detailed CM rules

Joint testing provides some scope for achieving economies of scale and flexibility in DSR testing, but some participants noted that the benefit was partially undermined by the requirement to demonstrate 100% capacity for joint tests as opposed to 90% for a single CMU in order to retain credit cover. In August 2017, Ofgem introduced a change to the CM rules to address this inconsistency.

Some participants reported that CM rules were biased towards baseload generators, rather than DSR (e.g. the requirement for 3 SPDs). But some generators reported the opposite bias, as the rules now allow DSR CMUs – but not existing generation CMUs - more flexibility to change components.

Our case study analysis found that the baseline methodology rules are ambiguous on how the average of the baseline demand samples is calculated, because of potential overlap in demand samples. The number of duplicates can vary between two and six of the 16 demand samples, depending on the date of the stress event, DSR test or SPD. The use of duplicate samples serves to skew the baseline average and could have a material impact on a test result.

Some participants felt that there is scope to improve the accuracy and hence fairness of the baselines**.** This can be achieved by statistical tests and/or adjustment mechanisms to account for independent variables, such as the weather, industrial production and the response to other market mechanisms. Day-of-delivery adjustment is commonly used in other markets as means to improve baseline accuracy while maintaining simplicity and transparency. This method is already used to adjust for responses to the balancing market, but only during stress events. Consideration could be given to extending the same or similar methodology to turn-down components to adjust for independent variables. But adding further complexity should be weighed against the short time-scales for National Grid and EMRS to process the data and the difficulties to date with adhering to regulatory timeframes.

Proposed rules for the second TA (Of12) require a full SPD retest if changes are made to a portfolio. Some participants reported that this risks discouraging portfolio maintenance. Aggregators with diverse portfolios using components from multiple clients are most likely to wish to carry out portfolio maintenance. However, these are the participants for whom SPDs are most disruptive and costly. This may incentivise them to hide reliability problems, as retesting is costly and risks damaging client relationships.

Impact of new technologies on DSR

Several TA participants commented that reductions in the cost of battery storage technology would significantly affect the future market for DSR. While some thought that battery storage would reduce the need for DSR, and ultimately make turn-down DSR redundant, others had a vision of storage becoming another asset in an organisation's flexibility portfolio, alongside generation and turn-down DSR.

Other external factors influencing DSR in the CM

Several TA participants commented that awareness of DSR within industry has grown significantly in the last year. This is likely to be attributable at least in part to the Power Responsive Campaign, combined with the opportunities for DSR in the TA, CM and other flexibility services. One participant commented that Power Responsive was effective at reaching larger organisations with energy managers, but not smaller organisations.

External factors such as proposed changes to embedded benefits by Ofgem and the review of flexibility services by National Grid will affect the future attraction of the CM by influencing the opportunities to stack CM revenues with revenues from Triad and other flexibility services.

9 Glossary and definitions

Term or acronym	Definition
Aggregator	An intermediary organisation that provides a service of collating capacity (from generation and/or DSR) for National Grid balancing services or the Capacity Market (CM), from a range of other organisations, in return for a share in the revenues generated.
Aggregator client	An organisation that contracts via an aggregator to access National Grid balancing services or the CM, rather than participating directly in these services.
Back-up generation	Generator (often diesel-powered) designed to be used if there is a power cut or problem with mains power. Usually located onsite 'behind the meter'.
Balancing services	System services contracted by National Grid. Those mentioned in this report comprise: 'Reserve services' that provide reserve capacity to balance electricity supply and demand (through generation or demand response). Examples include STOR and DSBR (see below). 'Frequency-related services' that provide very short-term changes in electricity demand or supply to help maintain the frequency of the grid for Great Britain (GB) at 50Hz. Examples include FFR and FCDM (see below).
Baseload generation	Electricity-generating equipment normally operated to serve loads on an around-the-clock basis.
Bayesian updating	The evaluation used an analysis tool called 'contribution tracing with Bayesian updating'. Bayesian updating refers to the specification of 'prior' probabilities for each hypothesis, and to the updating of these to 'posterior' probabilities, based on certain evidence tests. See Appendix 5 for more detail.
Capacity	The Capacity Market was established for the purpose of ensuring adequate capacity to meet the demands of consumers for the supply of electricity in Great Britain. Capacity can be in the form of electricity generation plant or reduction in demand for electricity.
Capacity Agreement	A capacity agreement comprises the rights and obligations accruing to a capacity provider under or by virtue of the CM Regulations and the Rules in relation to a particular capacity committed CMU and one or

	more delivery years.
Capacity Market (CM)	A series of auctions administered by National Grid, through which it procures future electricity capacity. The main auctions, known as 'T-4', are held annually 4 years ahead of the delivery year. Adjustments are made through annual 'T-1' auctions, one year ahead of the delivery year. The Transitional Arrangements involve two additional auctions that are designed to encourage growth in specific categories of capacity, to enable them to participate in the main CM in future.
Capex	Capital expenditure
CHP	Combined heat and power (a plant that produces heat as well as electricity).
CMN	Capacity Market Notice. The automatic warning issued by National Grid, warning that a stress event may occur in four hours' time. The criteria for issuing a CMN are automatic, and depend on the predicted balance between electricity supply and demand four hours ahead.
CMOs	Context-Mechanism-Outcome combinations. These are realist hypotheses about how the policy is expected to work, which are tested during the evaluation. See 'realist evaluation'.
CMU	Capacity Market Unit is a unit of electricity generation capacity or electricity demand reduction that participates in GB's CM. To pre-qualify for the first TA, a CMU had to be between 2 MW and 50 MW. For the second TA, the minimum threshold was 500 kW. A CMU may consist of a number of sites or components.
Component	A single site within a Capacity Market Unit (CMU). Some CMUs have only one component, while others have 20 or more. There is no lower limit on the capacity offered by a component, but there is an upper limit in that the sum of capacity offered by a TA CMU's components cannot exceed 50 MW.
Context	The circumstances which affect whether a policy 'works' and for whom. Consideration of 'context' forms an important part of realist approaches to evaluation.
Contribution analysis	Contribution analysis involves a structured process to develop and test a 'contribution story' (i.e. a coherent narrative explaining how a policy intervention appears to be influencing change, and assessing the likelihood that the intervention is contributing to observed results).
Contribution tracing	Contribution tracing involves the formulation and testing of competing hypotheses which could explain observed outcomes. The method involves explicit assumptions about the weight attached to different types of evidence, and aims to increase the transparency and

	replicability of qualitative analysis. See Appendix 5 for more detail.
Delivery year	The contractual year for delivery of CM obligations, which runs from 1st October of one calendar year through to 30th September of the following year.
De-rated capacity	Volume of generation or demand reduction capacity after a reduction to account for outage rates, maintenance down time and so on, which varies by technology type. National Grid publishes lists of standard de-rating factors by technology.
Distributed generation	Generation units which are connected to the distribution network, rather than the transmission network.
Distribution network	The electrical network that delivers electricity to the bulk of consumers (excluding a small number of consumers that are connected directly to the transmission network).
Direct participant	An organisation that participates in National Grid balancing services or the CM directly, rather than via an aggregator.
DNO	Distribution Network Operator. DNOs own and operate the distribution network of towers and cables that bring electricity from the national transmission network (see the National Grid) to homes and businesses.
DSBR	Demand-side balancing reserve (interim balancing service for winter 2015/16). One-to-one agreements between organisations and National Grid in which the organisation was paid to reduce demand at certain times. National Grid announced in August 2016 that they would not tender for DSBR in winter 2016/17.
DSR (Demand-side response)	DSR means the activity of reducing the metered volume of imported electricity of one or more customers below an established baseline, by means other than a permanent reduction in electricity use. See also 'Turn-down DSR' below. This report focuses on DSR by industrial and commercial rather than domestic consumers, as domestic DSR is much less well-developed in GB.
DSR Test	Test specified in CM rules, to demonstrate that a DSR CMU can reduce electricity usage by a given amount, relative to baseline demand. The test involves compilation of baseline data over a 6-week period, followed by collection of meter data by EMRS to confirm reduction below the baseline level at agreed times.
DTU	Demand Turn Up (the opposite of demand or load turn-down). Contracts with National Grid to make use of excess electricity generated by the distribution system (largely from solar power) when not otherwise

	needed.
Early Auction	An additional one-year ahead CM auction that has been introduced by BEIS. This was held in January 2017 and procured capacity for delivery in 2017/18. This auction cleared at £6.95/kW.
Electricity Settlements Company (ESC)	Government body set up to deal with paying capacity providers and recovering the costs from electricity suppliers.
Elexon	Organisation responsible for administering the GB electricity market Settlement and Balancing Code. Contracted by the ESC (see above).
Embedded benefits	Benefits negotiated between consumers and suppliers, when DSR or small-scale generation by electricity consumers helps suppliers to avoid network costs.
Embedded generation	Similar meaning to 'distributed generation'
Electricity Market Reform Settlements Limited (EMRS)	A wholly owned subsidiary of Elexon which the ESC contracts to settle CM payments and Contracts for Difference, and to collect and store metered data. Sometimes referred to as 'the settlements body'.
Enhanced Frequency Response (EFR)	A faster frequency response product tendered by National Grid in 2016, which requires organisations to interrupt their electricity supply within less than a second. Some of the service providers offer battery storage.
Fast Reserve	A service tendered monthly by National Grid that procures large blocks of reserve capacity (exceeding 50MW) that can respond within 2 minutes. Pump storage is currently the main provider of Fast Reserve.
FCDM	Frequency Control by Demand Management (similar to FFR). A bilateral agreement between an organisation and the National Grid, which requires the organisation to interrupt its supply for 30 minutes, at 2 seconds' notice.
Firm Frequency Response (FFR)	Firm Frequency Response. A monthly tendered service through which National Grid procures a few seconds or split-seconds of generation or demand reduction, at 30 seconds' notice, to support the 50Hz frequency at which the system operates.
Flexibility	Ofgem defines flexibility as 'modifying generation and/or consumption patterns in reaction to an external signal (such as a change in price) to provide a service within the energy system.'[29]

[29] Ofgem (2015), 'Making the electricity system more flexible and delivering the benefits for consumers.' *Ofgem Position paper*. Available at: https://www.ofgem.gov.uk/ofgem-publications/96959/flexibilitypositionpaperfinal-pdf Accessed 13 September 2016.

Flexible capacity	Electrical capacity (generation or load) that can offer flexibility to the electrical grid (see 'flexibility').
Frequency-related services	Services procured by National Grid to support the 50Hz frequency at which the system operates. These involve short-term changes to generation or demand at short notice, and usually require an automated response.
GB	Great Britain (the area covered by the electricity grid in England, Scotland and Wales).
Generative causation	Generative causation assessment methods involve a forensic examination of causality using strategic data collection and logic, rather than a probabilistic assessment of causation through statistical correlation. The basic explanatory structure in realism is that mechanism (M) acting in context (C) will generate outcome (O). These causal propositions (CMOs) are the starting point and end product of investigation in realist evaluation. See Appendix 4 for more.
Hassle costs	'Hassle costs' are the cost directly associated with TA participation. This could include marketing effort by aggregators, the cost of time spent on the TA application, auction and testing processes, and the cost of new metering or controls specifically required for the TA.
HLQ	High Level Evaluation Question – one of the main questions that BEIS has asked this evaluation to research.
Long-term STOR	Longer term version of STOR (see below), which was contracted by National Grid on a once-off basis but is now closed to new entrants. Holders of Long-term STOR contracts must declare that they will surrender these contracts if they obtain a capacity agreement for the same capacity.
Mechanism	A change in people's reasoning, brought about through the resources provided by a policy, which leads to a policy outcome. Identification of causal 'mechanisms', which operate in particular 'contexts', forms an important part of realist approaches to evaluation.
Missing money	In our analysis of costs and revenues associated with electrical capacity put forward for the TA, 'missing money' is defined to be the minimum revenue that a participant would require for their participation in the TA to break even.
NAO	National Audit Office
Net CONE	The 'net Cost of New Entry' is one of the parameters used to define the demand curve in a CM auction. It is set to reflect the estimated cost of marginal plant at the target capacity entering the auction.

National Grid	The National Grid runs Great Britain's electric high-voltage transmission network, is System Operator for the electricity system, commissions balancing services for the supply of flexible capacity and administers the GB Capacity Market. (See http://www2.nationalgrid.com).
Outcome	A change in the state of the world, brought about as a result of a policy or other influences. Realist approaches to evaluation attempt to identify the 'contexts' and 'mechanisms' that lead to a particular 'outcome'.
Proven DSR	A unit of DSR capacity that has passed the DSR test required to participate in the GB Capacity Market.
Realist contribution analysis	We have used the term 'realist contribution analysis' to describe contribution analysis that is undertaken in the context of a realist evaluation. Contribution analysis involves the specification of a theory of change, assessment of the evidence base, gathering of new evidence, theory testing and then refinement of the theory of change. While contribution analysis is often used to assess the 'average' contribution of an intervention, across a scheme as a whole, we have applied this method using a realist approach and have assessed the TA's contribution on a case by case basis.
Realist evaluation	A realist approach[30] to evaluation emphasises the importance of understanding not only whether a policy contributes to outcomes (which may be intended or unintended) but how, for whom and in what circumstances it contributes to these outcomes.
Realist hypotheses	Realist evaluation involves developing theories about programmes and policies. These theories involve the development of clear hypotheses about how, and for whom, programs might 'work'. The implementation of the programme, and the evaluation of it, then tests those hypotheses.
Realist synthesis	A realist synthesis is the synthesis of a wide range of evidence that seeks to identify underlying causal mechanisms and explore how they work under what conditions, answering the question "What works for whom under what circumstances?" rather than "What works?"
Red zone management	Consumers avoiding (or generators targeting) the times when distribution costs (or payments) are highest – i.e. the periods defined as 'red' or 'super red' in the peak demand traffic light system.
Reserve services	Contracts between National Grid and organisations that can provide capacity held in reserve, in the form of generation or DSR.
Satisfactory	CM participants are obliged to provide evidence of three half-hour settlement periods during the winter of a delivery year, on different days,

[30] Pawson and Tilley (1997) (op cit). Pawson (2006) (op cit).

performance days	in which they met their full capacity obligation.
Small-scale generation	For the purposes of the TA, generation units less than 50MW that are connected to the distribution grid.
STOR	Short-Term Operating Reserve - a reserve service run by National Grid through which organisations bid to provide generation or DSR to National Grid during peak demand periods (STOR windows). STOR is procured via tenders three times a year. A response time of at least 20 minutes is required.
Stress event	Period in which the electricity supply/demand balance is too tight (as determined by the System Operator's algorithms). Organisations holding capacity agreements are committed to provide capacity during stress events, or face penalties as set out in the CM rules.
Supplementary capacity market auction (also known as the 'Early Auction')	See Early Auction above
T-1	A one-year ahead CM auction, which will fine-tune the procurement of capacity in the main (T-4) CM auction for a given year. The first T-1 auction will be held early in 2018 and will secure agreements for the 2018/19 delivery year.
T-4	The main CM auction, held annually 4 years ahead of the delivery year. The first T-4 auction was held in 2014, procuring capacity to be delivered in 2018/19.
TA	Transitional Arrangements for DSR and small-scale distribution-connected generation – a pilot consisting of two one-year ahead CM auctions in 2016 and 2017 that are designed to encourage growth in specific categories of capacity, to enable them to participate in the main CM in future.
Transmission network	The high voltage power lines linking power stations to the distribution network. Some major electricity consumers are connected to the transmission network.
Triad avoidance	Consumers trying to reduce their electricity demand during three peak demand periods (or 'Triads'), in order to reduce their transmission charges. Transmission charges are based on demand during Triad periods. The Triad half hours are calculated from metered data (i.e. they are not known in advance) so Triad avoidance requires prediction of when the Triad periods might occur.
Triad targeting	Distributed generators trying to earn revenue by targeting generation at the Triad periods – the transmission charging methodology rewards

	them for doing so.
Turn-down DSR	Temporary reduction in electricity demand to avoid peak demand periods or to respond to National Grid instructions (sometimes called load reduction or curtailment). May also involve shifting electrical demand away from the peak demand period (sometimes called load shifting). This report focuses on DSR by industrial and commercial customers, as domestic DSR is much less well developed in GB.
Unproven DSR	A unit of DSR capacity that has not yet passed a DSR test, as specified by CM rules.