

C-MeX Pilot For PR19

REPORT PREPARED FOR



31 January 2019

Redacted version



PROC.01.0599

C O N T E N T S

<u>SECTION</u>	<u>PAGE</u>
Appendix ONE – Addressing Working Group Comments.....	1
1 BACKGROUND & OBJECTIVES.....	1
1.1 Background	1
1.2 Objectives	4
1.3 Key Areas of Investigation	5
1.4 Shadow Year	6
2 SUMMARY OF RECOMMENDATIONS.....	7
2.1 Customer Service Survey	7
2.2 Customer Experience Survey	8
2.3 Evaluation Scales And Sample Size	9
2.4 Calculating C-MeX Scores	9
2.5 Cross-Sector Comparison Gate	9
2.6 Outstanding Issues To Be Investigated In Shadow Year	10
3 Customer SERVICE SURVEY.....	11
3.1 Pilot Survey Methodology	11
3.2 Recommendations For Shadow Year	14
3.2.1 Methodology	14
3.2.2 Reasons For Not Pursuing SMS Survey Methodology	15
3.2.3 Online Survey Correction Factor.....	16
3.2.4 Sample Accuracy.....	18
3.2.5 Survey Frequency	19
3.2.6 The Questionnaire	20
3.2.8 Reallocation Of Interviews By Reason For Contact.....	23
3.2.9 Challenge Process	23
3.2.10 Hafren Dyfrdwy (HDD)	24
4 CUSTOMER EXPERIENCE SURVEY.....	25
4.1 Pilot Survey Methodology	25
4.2 Recommendations For Shadow Year	26
4.2.1 Methodology	26
4.2.2 Survey Frequency	27
4.2.3 The Questionnaire	27
5 CHOICE OF SCALE	28
6 SAMPLE SIZE AND COSTS	30
6.1 C-MeX Shadow Year Core Surveys	30
6.2 Validating The Correction Factor	33
7 CALCULATING A C-MEX SCORE.....	34
8 HIGHER PERFORMANCE PAYMENTS	39
8.1 Proposed Conditions For Higher Performance Payments	39
8.2 The Cross-Sector Comparison Gate	39

APPENDIX ONE – ADDRESSING WORKING GROUP COMMENTS

Please note that this document presents the results of a year of testing approaches to the C-MeX survey methodology. Any results in this document are to be read in the context that this is a pilot and are not necessarily a reflection of company performance.

Where information is sensitive it has been redacted for confidentiality reasons. Any redactions are marked with [X]

1 BACKGROUND & OBJECTIVES

1.1 Background

With a growing population, rising customer expectations and climate change affecting the water industry, Ofwat (the economic regulator of the water and wastewater sectors in England and Wales) is looking to develop newer and more flexible ways of regulating.

One element of Ofwat's role is to set the price, investment and service package customers receive through its price review every 5 years. To encourage water companies to provide an excellent level of customer service, the Service Incentive Mechanism (SIM) was introduced in 2010¹. This was (partially) a customer satisfaction measure and significant improvements in company performance have been seen over time as a result of the associated reputational and financial incentives. However, it is considered that there are limitations with SIM:-

- It is confined to the water industry and therefore does not act to encourage companies to achieve the higher levels of satisfaction seen in other sectors;
- SIM scores within the industry are converging and diminishing improvements at the higher end suggest limited effectiveness in supporting the top performing companies to go that step further;
- As companies are currently penalised for unwanted telephone contacts, they are discouraged from contacting customers, potentially constraining innovation, service improvement and customer engagement;
- The SIM survey is purely telephone based, however the way in which customers interact with service providers is constantly evolving with new technology;
- The cost and time constraints of a telephone methodology also limit the number of interviews achievable per water company per quarter; and
- It does not cover the customer service experience of developer services customers.

In July 2017, Ofwat publically consulted on an alternative set of mechanisms aimed at incentivising water companies to provide excellent customer service for residential customers (C-MeX) and developer services customers (D-MeX). Following further development through separate working groups including relevant stakeholders in the sector, the decision to replace SIM with C-MeX and D-MeX was made in December 2017. The key features as set out in Ofwat's December 2017 Final Methodology document² are outlined below.

¹ From March 2010 to April 2016, this survey was run and managed by McCallum Layton who changed their name to Allto Consulting in January 2017

²<https://www.ofwat.gov.uk/publication/delivering-water-2020-final-methodology-2019-price-review-appendix-3-customer-measure-experience-c-mex-developer-services-measure-experience-d-mex>

C-MeX – a single score based on combined results of 2 surveys:-

- Water company contactor satisfaction survey (Customer Service Survey)
- A general Customer Experience Survey (random selection of customers)

The Methodology stated that water companies will be ranked annually based on their C-MeX scores. The top three performers will receive a performance payment of up to 1.2% of residential retail revenues annually, whilst the poorest performers will receive an underperformance payment of up to 2.4% of residential revenues annually. In order to drive up customer satisfaction amongst the top performing companies a higher performance payment of up to 2.4% of residential retail revenue each year will be available providing a company:

- Performs at least on a par with a cross-sector threshold;
- Demonstrates satisfactory complaints performance (definition of satisfactory to be determined); and
- Is one of the top 3 companies by C-MeX score (if more than three companies meet the first two criteria).

The definition of 'complaint' is being widened from a definition largely based on the channel through which the complaint was made, to incorporate any complaint via social media and other channels where this can be attributed to a customer.

Companies will also need to offer at least 5 communication channels for receiving contacts and complaints. A negative adjustment to the C-MeX score is likely to apply if this is not realised. Unlike at present, C-MeX financial incentives will be applied on an annual basis.

Net Promoter Score (NPS) was also trialled in the C-MeX pilot.

D-MeX – again based on a single score:-

- Qualitative measure of water company performance based on a developer services customer satisfaction survey
- Quantitative measure of water company performance against a set of key metrics based on Water UK's existing metrics of service levels for developer services customers, and any future iterations of the metrics

The Methodology states that company performance will be ranked and published annually and financial performance payments of up to 2.5% of annual developer services revenue or penalties of up to 5% of annual developer services revenue will be applied to the best and worst performers.

Navigator Research, and its sub-contractor Allto Consulting, were commissioned to design and run pilots of both mechanisms to test their efficacy and refine the methodology in advance of them becoming fully

operational in April 2020. This document outlines the results and recommendations arising from the pilot C-MeX research.

1.2 **Objectives**

The overall objective of the project, as stated by Ofwat, was to:

1. Test all aspects of C-MeX so the incentives meet the following success criteria: -
 - encourages companies to improve customer experience and innovate;
 - simple and meaningful for companies and customers;
 - proportionate;
 - practical to implement;
 - measures performance across companies consistently, reliably and fairly; and
 - reflects customer behaviour changes and market changes.

To develop the mechanisms to the point where they can be run in shadow form (which is intended to be a dry run) in 2019-20, and have financial incentives applied on an annual basis from April 2020 onwards. This involved designing and running at least two waves of pilot customer surveys for C-MeX that are consistent with the design decisions of each incentive outlined in the methodology statement 'Delivering Water 2020: our methodology for the 2019 price review' (sections 3.1 and 4.1).

A further subsidiary objective was to:

2. Provide advice and make evidence-based recommendations on some of the outstanding elements of the C-MeX design not covered by the decisions published in the PR19 methodology statement, including:
 - whether NPS should form part of the financial incentive (section 7);
 - the weightings of the customer service and customer experience surveys within the C-MeX calculation (section 7);
 - the most appropriate source of data for the cross-sector threshold in C-MeX (section 8); and
 - how to adapt C-MeX, if necessary, based on the results of the C-MeX pilot (sections 3.2 and 4.2).

1.3 Key Areas of Investigation

For the purposes of the pilot, various approaches were trialled in Wave 1 and Wave 2 to determine the most appropriate ones to take forward to the shadow year. These are outlined below:-

1.3.1 C-MeX Pilot Wave 1 (May-June 2018)

Customer Service Survey

- Survey approach in line with how customers contacted their water company – telephone survey for telephone/letter contact and an online survey for customers making digital contact
- Amongst those customers contacting by mobile – comparing an SMS survey to the telephone survey
- Customer contact data for all companies provided weekly over a 3 week period to mitigate against one-off events and ensure sufficient sample

Customer Experience Survey

- Focus on company providing clean water service
- Random Digit Dialling telephone approach using postcode areas to determine water provider
- Age and gender quotas in line with customer base

Both Surveys

- Boosting water company samples (optional by water company) to improve accuracy of results
- Alternative scoring scales – in order to determine the ability of different scales to differentiate between water companies and give a greater spread of results (the limited spread of results is a drawback of SIM), and establish their usefulness as a comparison versus UKCSI (UK Customer Satisfaction Index, as measured by the Institute of Customer Service), we randomised the scale used across the sample, so that half of each water company's responses to the customer satisfaction question were rated on a verbal 1-5 scale and half on a numeric 0-10 scale

1.3.2 C-MeX Pilot Wave 2 (September – October 2018)

Customer Service Survey

- Investigate why online survey results are significantly more negative than telephone survey ones by carrying out some telephone interviews with digital contactors to determine whether the difference is caused by methodology or poorer performance of digital channels
- Examine the current Billing/Clean/Waste weighting and establish whether an alternative weighting system might be fairer

Customer Experience Survey

- Investigate any differences in results by methodology (face-to-face versus telephone)

Both Surveys

- Survey frequency for the shadow year
- An appropriate sample size for the shadow year taking into account the need for a robust sample and cost

1.4 Shadow Year

It is intended to run the surveys in shadow form from April 2019 – March 2020, to give companies time to make any systems changes needed before the financial incentives are applied from April 2020 onwards. Another purpose of the Shadow Year is to allow Ofwat to monitor the operation of C-MeX (and D-MeX) incentives and make adjustments which it considers appropriate for the period 2020-2025.

Our recommendations outlined in this report apply to this shadow year, during which any outstanding issues can be tested and discussed with stakeholders. Those companies with reputational incentives for SIM in 2019-20 will be able to use the Customer Service element of C-MeX (excluding online interviews) as a proxy for the qualitative component of SIM. Ofwat's PR19 Final Methodology stated that final C-MeX guidance will be published by March 2020.

2 **SUMMARY OF RECOMMENDATIONS**

In summary, our recommendations for the C-MeX shadow year are as follows:-

2.1 **Customer Service Survey**

- A mixed methodology approach in which as many online interviews as possible are achieved by emailing up to 10,000 digital contactors per water company (sample permitting) and conducting telephone interviews with 200 non-digital contactors per company as indicated in the table below:-

Proposed Unweighted & Weighted Numbers Of Interviews For C-MeX Customer Service Survey				
	Unweighted		Weighted	
	WASC³	WOC⁴	WASC	WOC
Billing	67	100	100	100
Clean Water	67	100	50	100
Waste Water	66	n/a	50	n/a
Total	200	200	200	200

- Online and telephone survey results should be weighted to be reflective of the relative proportions contacting using digital/non-digital channels respectively within the supplied sample of billing/operations contactors.⁵
- In order to prevent any perverse incentive for companies not to expand digital contact methods, an upwards 'correction factor' should be applied to results from the online survey to take account of the fact they are typically more negative due to the self-selecting nature of the methodology (participation in an online survey is completely led by the consumer, whereas in a telephone survey the interviewer is able to influence participation). This correction factor should be in the region of 5%, to be verified in the shadow C-MeX year (section 3.2.3).
- Sample quality, in particular coverage of email addresses for digital contactors, should be monitored going forward to maximise the representativeness of customer samples provided by water companies. Despite improvements between the Wave 1 and Wave 2 pilots, some water companies have an unacceptably high level of digital contactors without email addresses.
- Fieldwork for the survey should be carried out on a monthly basis with one week each month selected in advance by the appointed agent and Ofwat as being the sampling week (water companies would not be party to this information). Under this approach, water companies would be informed on the Monday morning following the sampling week that they will need to provide the sample by 5pm the

³ WASC = Water And Sewerage Company

⁴ WOC = Water Only Company

⁵ We currently define a contactor as an identifiable customer who makes direct inbound contact with their water company.

following day. Smaller WOCs would be required to submit two weeks' of operations data to ensure interview numbers are achieved. We recommend that any data supplied to water companies should be done each month following fieldwork but results should be rolled up and reported in league table format on a quarterly basis.

- Despite accounting for over 90% of all contacts, we recommend the weighting of billing results within the overall sample should be 50%, as in SIM, to ensure water companies focus efforts to improve customer service across both the billing and operational parts of their business.
- Two alternatives have been proposed in terms of the weighting of Clean and Waste results within Operations for WASCs – 33% Clean/17% Waste to reflect the relative proportions contacting their water company with these issues (favoured by WOCs as being fairer for them, as Waste scores are typically higher than Clean ones) and 25% Clean/25% Waste as in SIM (favoured by most WASCs in order to ensure Waste customer satisfaction does not suffer as focus shifts to performance on Clean Water on both the Customer Service and Customer Experience surveys). Either option is viable, therefore we recommend that Ofwat considers which they find most appropriate for C-MeX, balancing perceived fairness and ensuring water companies focus efforts on all areas of their business. Results within each contact type should also be weighted in proportion to the balance of digital and non-digital contact received from each water company for that contact type.
- Where there exists a discrepancy between water company and respondent reason for contact after prompting, respondents should be re-allocated to the respondent's stated contact type on completion of interviewing to ensure more accurate reporting.
- Hafren Dyfrdwy (HDD) is a new WASC with very few waste water contacts itself, but who use the same outsourced provider for waste water services as Severn Trent Water. With insufficient waste water interviews for a robust sample we recommend that Ofwat considers the option of interviews being carried out with waste water customers from the combined Severn Trent/HDD Waste Water service. If adopted, waste water results would be shared between the two companies in the same way as Bristol Water and Wessex Water share billing results.

2.2 Customer Experience Survey

- A predominantly telephone-based survey using a Random Digit Dial (RDD) sampling approach to achieve a broadly representative sample, supplemented by targeted electoral roll sampling and face-to-face interviews to ensure sufficient interviews are achieved with younger age-groups.
- There should be an equal number of face-to-face interviews carried out for each water company amongst 18-44 year olds, representing 20% of all interviews (i.e. 40 interviews, 30 with 18-29s and 10 with 30-44s). This will ensure fairness across water companies. Interviews should be carried out in 2 locations for each water company each quarter and interviewing locations should be rotated across the water

company region. Samples should therefore reflect the shape of the population.

- Water company serving the customer should be verified using customer postcode
- Each water company should supply the appointed agent with full postcode information for all clean water customers on an annual basis.
- All water service users aged 18+ would be eligible to take part in the survey, providing they are aware of their water company either spontaneously or on prompting; respondents do not have to be the bill payer.
- Quotas should be set on gender and age (18-29, 30-59, 60+) to be reflective of each company's demographic profile based on 2011 Census data for the local authorities served by each water company. These will ensure an appropriate spread of people consistently for each water company whilst being simple, proportionate, practical to implement and cost-effective given the sampling methodology (random digit dialling, so no prior knowledge of age).

The survey should be carried out on a continuous basis over the course of each quarter.

2.3 Evaluation Scales And Sample Size

- We recommend using a 0-10 rating scale for customer satisfaction to provide greater differentiation in scores between water companies
- We recommend a sample size of 200 telephone interviews per survey per company per quarter plus as many online interviews as can be achieved on the Customer Service survey from sending up to 10,000 email invitations per company to those customers contacting by digital means. We feel this is the optimum sample size in terms of balancing robustness and cost.

2.4 Calculating C-MeX Scores

- On the basis of transparency and simplicity, our recommendation is a 50%/50% weighting of the Customer Service and Customer Experience customer satisfaction results.

2.5 Cross-Sector Comparison Gate

- Our recommendation is that the cross-sector threshold calculation proposed by Ofwat is adopted i.e.

$$\text{C-MeX ASUQ} = \text{C-MeX Mean} + (\text{UKCSI ASUQ} - \text{UKCSI Mean}) / \text{UKCSI SD} * \text{C-MeX SD}$$

Where ASUQ = All Sector Upper Quartile score and SD = Standard Deviation – see full explanation in Section 8.2

- This results in a threshold score of 86.6 for the combined Wave 1 and 2 pilot – the highest scoring water company currently achieves a C-MeX score of 84.9, therefore this feels to be a stretching threshold without being unattainable.

2.6 Outstanding Issues To Be Investigated In Shadow Year

Customer Service Survey

- The quality of samples supplied by water companies in terms of accurate contact details (particularly email addresses)
- The rate at which digital contact is replacing non-digital contact
- The effectiveness of monthly sample provision and interviewing
- The effectiveness of a shared waste interview approach for Severn Trent and Hafren Dyfrdwy
- The exact correction factor to be used for online survey results (this has not been costed for in Section 6)

Customer Experience Survey

- The effectiveness of postcode verification of consumers

C-MeX

- The stability of the overall C-MeX measure
- The effectiveness of the Cross-Sector threshold in driving improvements

3 **CUSTOMER SERVICE SURVEY**

3.1 **Pilot Survey Methodology**

Ofwat adopted a mixed methodology approach to be tested in the pilot in which customers were interviewed by the method by which they contacted their water company – i.e. an online survey for any customers contacting their water company by digital means and a telephone survey for those contacting by telephone or in writing (through the post). In Wave 1, the opportunity was also taken to test an SMS methodology amongst customers contacting by mobile phone.

Table 1 shows the methodologies used for different types of contactors in each wave and the target number of interviews for the telephone methods. In both Wave 1 and Wave 2 of the pilot, water companies were given the option of funding a boost to their sample sizes on the Customer Service CATI survey, therefore in the table below, 'Core' was the required sample size in the absence of boosting and 'Optional Boost' the required sample size for a boosted sample.

Table 1 – Customer Service Survey: Interviews Per Water Company				
Research Method				
Type of contact	Telephone	Online	SMS	Telephone with Digital Contactors
Wave	Non-Digital Both Waves	Digital Both Waves	Mobile Wave 1 Only	Digital Wave 2 Only
Billing				
Core	50			
Optional Boost	150			
Total	200			
Operations*				
Core	100	As many as possible from up to 10,000 records	As many as possible from up to 500 mobile numbers	25 per water company
Optional Boost	100			
Total	200			
Total				
Core	150			
Optional Boost	250			
Total	400			

* Split 50% Clean Water and 50% Waste Water contactors for WASCs

CATI (Telephone) Survey – Both Waves

The majority of water companies' contacts were by telephone. Most companies chose to boost their samples to 400 interviews, split evenly between billing contacts and operational contacts except for:-

- Northumbrian Water, Portsmouth Water and South East Water in Wave One
- Hafren Dyfrdwy, Portsmouth Water, South West Water, South Staffs Water and South East Water in Wave Two

Each of these companies achieved a core sample of 150 interviews, split 50 billing contacts and 100 operational contacts.

No other quotas were applied. In order to be eligible respondents had to recall contacting their water company.

Online Survey – Both Waves

An online survey was adopted as the research tool for any type of digital contact from customers to their water company, such as email, online transactions, webforms, live chat, social media etc. Customers using multiple channels i.e. phone and digital were interviewed through the channel used for their most recent contact.

An email invitation containing a link to the survey was sent out to up to 10,000 customers per water company who contacted by digital means (not all companies had sufficient digital contacts to provide this level of sample).

In Wave One, 115,346 customers were approached to take part in the online survey; 4,652 responded, a 4.0% response rate. In Wave Two, 125,458 customers were approached to take part in the online survey; but only 3,680 responded, a 2.9% response rate. The reduction in response rate could be a result of changes to the invitation email wording and the inclusion of an opt-out link between Wave 1 and Wave 2 as a result of the introduction of GDPR legislation in May 2018.

SMS Survey – Wave One

For the Wave 1 pilot we trialled a very short (2 question) SMS survey to a sample of customers contacting by mobile in order to test the representativeness and reliability of this as a survey method.

We sent out an SMS survey to a total of 6,483 customers who had contacted by mobile across all water companies (we were aiming to send up to 500 interviews per company but some companies lacked sufficient mobile sample to support achieving these numbers). The outcome of these responses was as follows:-

- 763 people (12%) responded with 'STOP' meaning they didn't want to take the survey further
- 398 (6%) responded to continue

- 353 people (5%) completed the first question
- 342 people (5%) answered the second question

The number of responses per water company varied from 6 for Hafren Dyfrdwy (previously Dee Valley) to 35 for Northumbrian Water, but averaged 21 per water company.

SMS survey results were not included in the overall results as only two questions were asked and there was no respondent confirmation of reason for contact.

Telephone Survey with Digital Contactors – Wave Two

In Wave 1, Contactors were interviewed using the same channel they used to contact their water company, i.e. an online survey for digital contactors. However, a lower satisfaction score amongst digital contactors was observed and it was not apparent if this was a result of poor query handling through digital channels or the self-selecting nature of an online methodology.

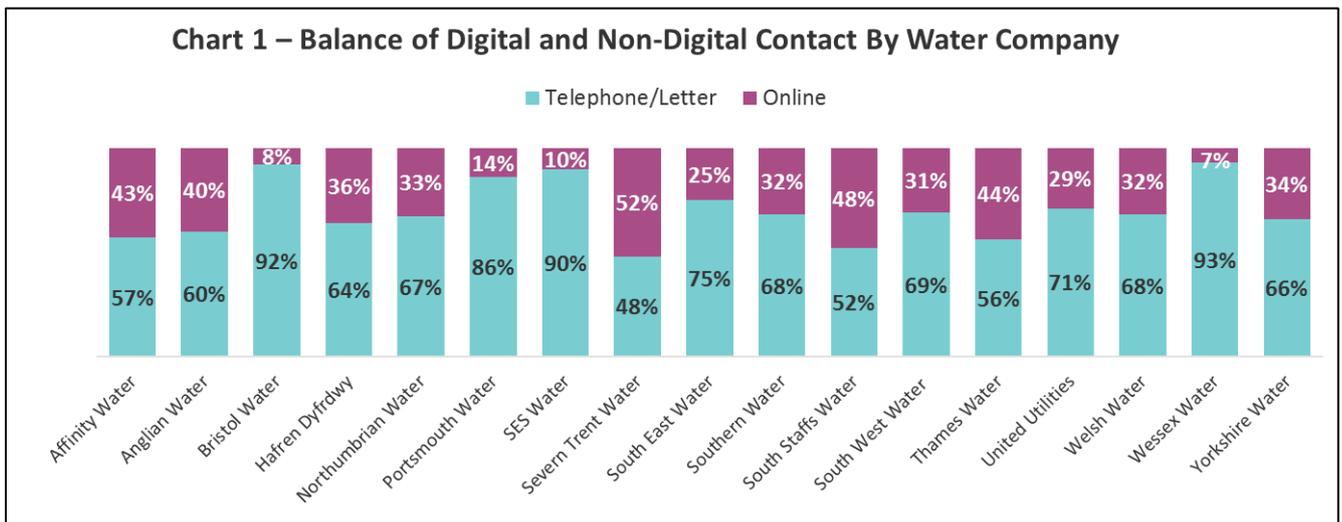
As such, in Wave 2, 425 additional telephone interviews were carried out with digital contactors across the industry in order to directly compare the online and telephone methodologies and determine the cause of the lower scores. The survey results for these respondents were not included in overall results

3.2 Recommendations For Shadow Year

3.2.1 Methodology

We recommend a mixed methodology approach to the Customer Service survey as in the pilot, using telephone interviewing for customers contacting their water company by non-digital means and an online survey for those contacting their water company by digital means.

This approach should ensure that a robust sample of telephone interviews is achieved with those who contacted their water company by non-digital means (given that the majority of contacts with all water companies but one are through non-digital channels and these customers may not have access to, or be reluctant to use digital channels to respond), whilst allowing for the fact that digital contact now forms a substantial and growing proportion of most companies' contacts (see Chart 1).



On the telephone survey we recommend carrying out an equal number of interviews with billing, clean and waste contactors for WASCs and an equal number of interviews with billing and clean contactors for WOCs. Results should then be weighted to be more representative of the make-up of water company contacts.

Based on a sample size of 200 per company per quarter and weightings of 50% billing, 50% operations (with a 25%/25% split of clean and waste water contacts for WASCs as per SIM), each water company would aim to achieve the following number of telephone interviews for each contact type each quarter:-

Table 2 – Proposed Unweighted & Weighted Numbers Of Telephone Interviews For C-MeX Customer Service Survey				
	Unweighted		Weighted	
	WASC	WOC	WASC	WOC
Billing	67	100	100	100
Clean Water	67	100	50	100
Waste Water	66	n/a	50	
Total	200	200	200	200

As mentioned in Section 3.1, online survey response rates are low (c. 3-4%), which means that a large number of survey invitations need to be emailed out to achieve robust numbers of interviews. However, as online interviews captured in this way are relatively low cost (assuming no coding of open-ended questions into themes), an online survey provides a cost-effective method of substantially increasing the robustness of annual sample sizes and thereby addressing water company concerns about sufficient robustness.

The achievable sample size is limited by the number of customers contacting their water company through digital means during the sampling week(s), and this does vary by company – the smallest water company, Portsmouth Water, is likely to achieve a maximum of only 200 interviews a year on present numbers of digital contactors (approximately 230 useable records per week), whereas the bigger WASCs could expect to achieve up to 1,600 interviews per year by emailing 10,000 contactors a quarter with the survey.

In total, six water companies did not supply sufficient digital contactors with useable email addresses to send out 10,000 invitations in each wave of the pilot; improved coverage of email addresses will mitigate this to some extent (see Section 3.2.4).

There is discussion of alternative weighting of different contact types in Section 3.2.7 and of alternative sample sizes in Section 6.

3.2.2 Reasons For Not Pursuing SMS Survey Methodology

The SMS survey approach is not recommended to be adopted as a method of data collection for C-MeX going forward for a number of reasons:-

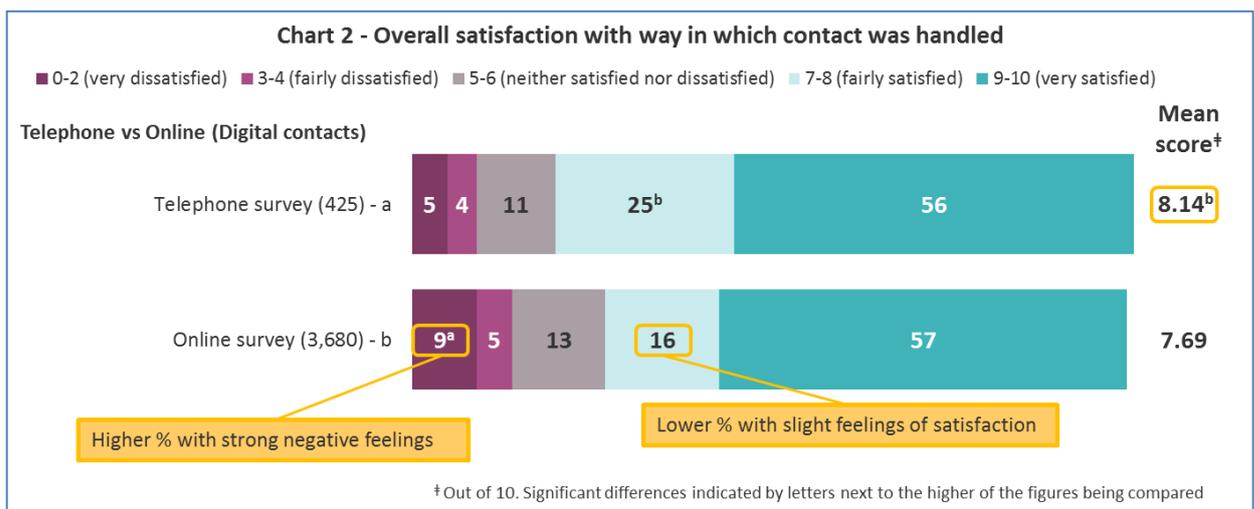
- The survey has to be very short for respondents to be willing to take part/complete the survey
- It can only be sent out to those customers for whom the water company holds a mobile number – this varies by water company from 53% to 79% of contactors, averaging 66%. If this was the only interviewing methodology, on average a third of contactors could not be surveyed (likely being older/more vulnerable people with a landline only), potentially resulting in a biased sample

- There has to be an assumption that customers are thinking about the correct occasion on which they contacted the water company - unlike telephone or online interviewing there is no opportunity to collect the reason for contact the customer is thinking about
- The response rate is much lower than telephone interviewing (5.0% versus 13.4%), and like the online survey, results are significantly less positive than telephone results due to the self-selecting nature of the method, therefore the sample achieved is less representative of all water company contacts.

3.2.3 Online Survey Correction Factor

In Wave 1, Contactors were interviewed using the same channel they used to contact their water company, i.e. an online survey for digital contactors. However, a significantly lower satisfaction score amongst digital contactors was observed and it was not apparent if this was a result of poor query handling through digital channels or the self-selecting nature of an online methodology. Moreover, online response rates (2.9%-4.0%) were substantially lower than for telephone surveys, where interviews were achieved with 13.4% of those contacted. A low response rate can give rise to sampling bias if the non-response is unequal among those approached, in terms of outcome eg if customers with strong negative feelings are more likely to respond than those who are ambivalent, suggesting the online survey may be less representative of the contactor population than the telephone one in terms of the service they received.

As such, in Wave 2, 425 additional telephone interviews were carried out with digital contactors across the industry in order to directly compare the online and telephone methodologies and determine the cause of the lower scores. Results are shown on the following chart.



The chart shows that those completing online surveys were significantly less positive than those interviewed over the phone. From the distribution of scores, there were significantly fewer respondents who were reasonably satisfied (scores of 7-8) in the online survey versus the telephone one, and significantly more giving scores of 0-2. This indicates that those who felt

their contact was handled 'ok' were less likely to bother to respond to the survey online, whilst those who felt it was particularly poor were more likely to be galvanised into responding.

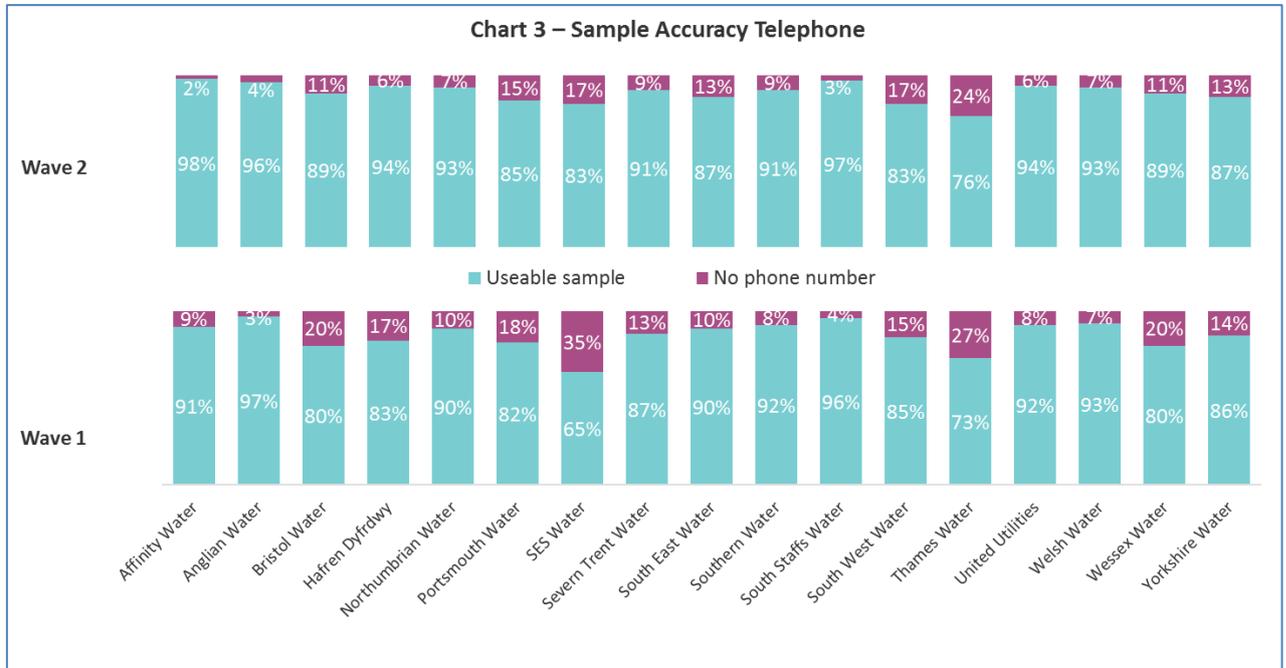
The result of this is that companies with higher proportions of digital contact (and therefore higher numbers of online surveys) will be penalised in their C-MeX scores versus those with lower proportions. This might act as a deterrent to companies directing customers to digital contact channels. This is obviously contrary to Ofwat's aims of making water companies as accessible as possible to their customers by offering a range of contact methods. As such, we recommend applying a correction factor to the online results to bring them in line with the CATI scores and remove any potential bias.

As the telephone results amongst digital contactors are 5.9% better than online results, this suggests the correction factor should be around 5.9%. At the December C-MeX Working Group Workshop, one water company offered the use of their own telephone survey tracking data with digital contactors (based on 3,402 interviews over the previous 18 months) as a comparison to their C-MeX online survey results (based on 752 interviews) in order to validate this figure. In this case the ratio of scores came out as being slightly lower, at 4.3%. As this company's results are based on a more robust sample of telephone interviews, we would expect the true correction factor to be closer to 4.3% than 5.9%. This is supported by a direct comparison between all online interviews and all telephone interviews in each wave, which indicates a difference of 4.4% in Wave 1 and 5.0% in Wave 2.

We recommend a large-scale test during the shadow year, using a telephone methodology to achieve interviews with a larger sample of digital contactors (~1,000) and comparing this to digital contactor interviewers achieved online as a verification of the factor. Differences by company would also be analysed. We don't believe the factor will change substantially over time as it is a consequence of the difference in methodologies, rather than a difference in attitudes or performance, therefore we feel a robust survey in the shadow year to verify the factor will be sufficient to cover the period 2020-2025.

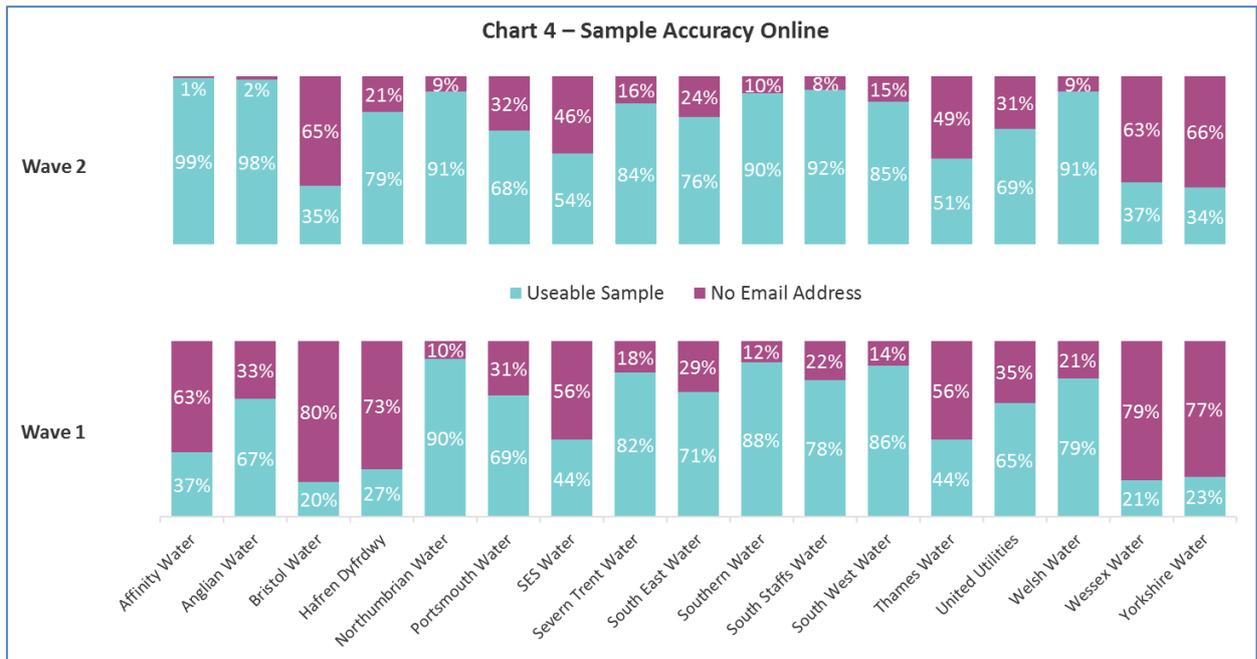
3.2.4 Sample Accuracy

The proportion of useable non-digital contactor sample (i.e. telephone, letter and visit) provided by water companies in Wave 1 and Wave 2 are shown in Chart 3 below.



A minimum of 76% of customer telephone numbers were useable amongst non-digital contactors of each company in Wave 2, and most companies had a substantially higher proportion than this, indicating the samples used for interviewing are likely to be highly representative of water company non-digital contactors. There was typically an improvement in the accuracy of the telephone contact details provided by water companies from Wave 1 to Wave 2.

Chart 4 shows the proportion of useable digital contactor sample (e.g. email, social media, webchat, webform etc.) provided by water companies in Wave 1 and Wave 2.



Sample accuracy for digital contactors (in terms of email coverage) tended to be worse than for non-digital contactors (in terms of telephone number coverage), and despite big improvements from Wave 1 to Wave 2, several companies had a large proportion of digital contactors for which they did not have digital contact details, which reduces the representativeness of their samples. It is important for all companies to have correct email addresses for as many of their customers as possible in order to ensure results are a fair reflection of the entirety of their digital offering.

3.2.5 Survey Frequency

For the shadow year we recommend a monthly Customer Service survey whereby the appointed agent, in conjunction with Ofwat, will select a random 'C-MeX Sampling Week' each month, in advance. Water companies will then be notified by 10am on the Monday following the C-MeX Sampling Week that they will be required to provide all records of customers contacting the company in that week to the appointed agent by the end of the following day (i.e. 5pm on the Tuesday). Smaller water companies will be asked to provide two weeks of operations data to ensure sufficient interviews can be achieved with operations contactors. We recommend that results are rolled up and reported in league table format on a quarterly basis.

This approach will help to ensure water companies provide consistent good service as they will be surveyed broadly speaking one week in every four and any of those four weeks could be the sampling week. Moreover, it will mitigate against one-off events (e.g. sewer flooding or hosepipe bans) which are less likely to skew survey results for a particular company or companies if sampling is spread over 3 weeks in the quarter (or 6 for the smaller WOCs) rather than 1 week as at present for SIM.

As the number of quarterly interviews should be spread over three months, the total number of interviews to be achieved each month will be less than SIM, therefore all interviews should be completed within 2 weeks of the customer contacting their water company, which will mean the contact is fresh in their mind.

Whilst league tables would be released each quarter, we recommend data is supplied to water companies on completion of interviewing each month.

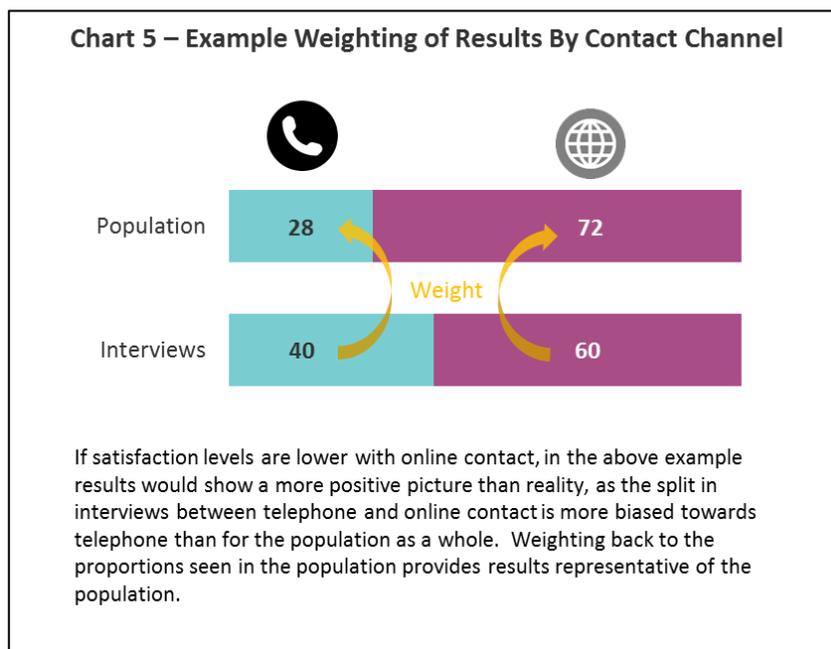
3.2.6 The Questionnaire

Going forward, we recommend that Q3 is removed from the Customer Service questionnaire (whether customers took part in other research with their water company in the last 3 months) as this was included to prevent overlap with the SIM survey whilst both were undertaken in a similar time frame and will no longer be required in the shadow year. With this exception the questionnaire should remain the same as for the pilot survey Wave 2, at 7 minutes in length.

3.2.7 Weighting Of Results

Weighting by Contact Channel

Due to the self-completion elements of the research, and the differing numbers of available customers contacting online, the total number of interviews achieved (CATI and online) will vary across companies. In order that results for each water company are representative of customer contact with that company we recommend weighting results within billing and within operations, based on the relative proportions of online and telephone/written contact within the samples sent by each company.



For example, if a company provides a sample composed of 72% online customer contacts and 28% telephone for its billing function and achieves 150 online interviews and 100 telephone interviews (i.e. a 60%/40% split of online and telephone interviews), we would weight online up so that it accounts for 72% of the overall company satisfaction score for billing (i.e. 180 interviews) and weight telephone down so that it represents 28% of the overall company satisfaction score for billing (i.e. 70 interviews). The same would be done within Operations and then overall billing, and operations results will be weighted according to the desired proportions in the final sample. This is shown in the table below.

	Achieved no. interviews	Achieved % of interviews	Required % of interviews*	Weighted no. interviews	Final weighting**
Billing					
Online	150	60%	72%	180	137
Telephone	100	40%	28%	70	53
Total	250	100%	100%	250	190
Operations					
Online	30	23%	5%	7	10
Telephone	100	77%	95%	123	180
Total	130	100%	100%	130	190

* From provided sample

** Assuming 50% billing, 50% operations weighting

Under SIM, results are weighted to 50% billing, 25% clean water and 25% waste water for WASCs; and 50% billing and 50% operations for WOCs. The composition of water company submissions across both pilot waves was analysed to determine the appropriateness of these weightings for C-MeX. The industry level split of billing, clean and waste contacts in C-MeX pilot Wave 1 and Wave 2 is shown in the table below:-

Table 3 – Sample Composition: Reason For Contact			
	Wave 1	Wave 2	Overall Pilot
WASCs			
Billing	90%	92%	91%
Clean Water	7%	6%	7%
Waste Water	3%	3%	3%
WOCs			
Billing	93%	94%	94%
Clean Water	7%	6%	6%
Base: All water companies			

It can be seen that the overwhelming majority of customer contacts in the water industry are billing related. Weighting final results to be representative of the mix of reasons for contact is not desirable as it would mean that Customer Service scores would be almost exclusively driven by billing results, with operational results having only a marginal impact on the overall score. This may well then lead to a decline in operational customer service levels as companies focus on providing an excellent billing service.

Retaining a 50% billing/50% operations weighting for C-MeX should mean that companies continue to make it important to provide excellent customer service in both business areas equally, therefore this is our recommendation.

However, one potential change that could be made to the weighting is the balance of clean water and waste water results within operations for WASCs. As can be seen from the table, clean water contacts are over twice as common as waste water contacts, therefore there is an argument for applying a weighting of 50% billing/33% clean/17% waste (rather than 50%/25%/25%). Such a weighting is more reflective of the actual balance of operational contacts and would address WOC concerns about an equal split being unfair – the argument being that waste water results are typically more positive than clean water ones, so if waste water is over-represented in the final results, WASC scores are inflated.

The counter argument to adopting the alternative weighting is that reducing the importance of waste water in the Customer Service survey, coupled with the Customer Experience survey's focus on water users only, may result in excellent waste water contact handling becoming less of a priority for WASCs (although it will still represent almost a fifth of their Customer Service score).

The impact on the Wave 2 pilot league table of adopting the revised weighting is shown below:-

Wave 2 Weighting:	WASCs 50% Billing/ 25% Clean/ 25% Waste	WASCs 50% Billing/ 33% Clean/ 17% Waste	Rank
Industry	CSAT	CSAT	
Industry	7.88	7.85	
Portsmouth Water	8.63	8.63	1
Anglian Water	8.49	8.44	2
South West Water	8.41	8.38	3
Northumbrian Water	8.40	8.37	4
Wessex Water	8.39	8.34	5
United Utilities	8.22	8.15	6
South Staffs Water	8.06	8.06	7
Welsh Water	8.06	7.93	7/8
Yorkshire Water	7.82	7.79	9
Bristol Water	7.74	7.75	10
Severn Trent Water	7.73	7.64	11
South East Water	7.59	7.59	12
Affinity Water	7.43	7.43	13
Southern Water	7.43	7.39	13/14

Thames Water	7.25	7.16	15
SES Water	7.16	7.16	16/15
Hafren Dyfrdwy	7.01	6.98	17

From the table, all WASC scores fall slightly under the alternative weighting system, although there is minimal impact on actual rankings – two WASCs drop a ranking and one WOC improves. Applying the alternative weighting system in Wave 1 benefitted one (mid table) WOC; the only other shifts were WASCs exchanging rankings.

Either of these weighting options is viable, therefore we recommend that Ofwat considers which they find most appropriate for C-MeX, balancing perceived fairness and ensuring water companies focus efforts on all areas of their business.

3.2.8 Reallocation Of Interviews By Reason For Contact

To ensure respondents were talking about the correct contact, a check was built into the Customer Service telephone survey questionnaire such that if a discrepancy arose, the customer was prompted with the reason given by the water company to see if they also had contact about this issue and if so, interviewed on that issue; otherwise they were interviewed about their claimed reason for contact. If respondents were clearly confused about who they spoke to and/or appeared to be talking about an organisation other than their water company, the interview was terminated.

Conducting such a check requires an interviewer, as the terminology used by water companies may often not be straightforward/easy to understand or different to a customer's perception (eg a customer may ring up about a high bill but the water company may class this as a leak, which has caused the high bill). Whilst an interviewer can often help to interpret the water company terminology to check it matches with the customer's stated reason; or make the decision to continue with the interview using the customer's stated reason if the reason provided by the company is not clear, this is not available for those self-completing the online survey. In the case of the online survey, we just asked respondents to record their reason for contact, then continued with the interview.

During the C-MeX pilot, 5% of interviews in Wave 1 and 4% in Wave 2 had a discrepancy between water company and respondent reason for contact. Following Wave 2 these were re-allocated to the contact type provided by the respondent on completion of interviewing to ensure more accurate reporting. We recommend this is carried out going forward into the shadow C-MeX year.

3.2.9 Challenge Process

Within the SIM survey there is a process in place for water companies to query or challenge their data to ensure the robustness and accuracy of the results, subject to data protection rules. This has not been part of the pilot phase.

We recommend that Ofwat considers whether it is appropriate to adopt a similar process for the C-MeX shadow year.

3.2.10 Hafren Dyfrdwy (HDD)

Hafren Dyfrdwy became a new WASC in July 2018, prior to the Wave 2 pilot survey. Whilst the first data submission week of the Wave 2 pilot realised around 900 customers contacting the company regarding waste water issues, this appeared to be atypical, with numbers dropping to an average of 7 customers per week for subsequent weeks. On the basis of 7 contacts a week, we would expect to achieve only around 8-10 interviews per quarter with contactors with waste water issues (assuming 2 weeks' of operational contactors are provided each month). This is insufficiently robust to treat Hafren Dyfrdwy in the same way as other WASCs.

There are two potential solutions to this:-

- i) Treat the company as a WOC and combine clean and waste water results to produce a single operations score
- ii) Combine waste results for Severn Trent Water and Hafren Dyfrdwy to produce one waste water score for both companies, given that Severn Trent Water have an outsourced model and the same provider services both Severn Trent Water and Hafren Dyfrdwy waste water customers

If option i is adopted, there is a risk that the waste water service provider (and, potentially, internal teams) will deprioritise Hafren Dyfrdwy waste water service. Option ii would ensure waste water service levels remain important across both companies. There is a precedent for option ii, with the approach to the combined billing service for Bristol Water and Wessex Water – billing interviews are carried out with Bristol & Wessex Billing Services Limited contactors and shared across the two companies. We recommend that Ofwat considers adopting Option ii.

4 CUSTOMER EXPERIENCE SURVEY

4.1 Pilot Survey Methodology

A telephone survey was carried out initially using a Random Digit Dial (RDD) landline and mobile sampling approach to obtain a representative sample of customers for each water company. Each water company provided details of the postcode sectors it operates in for clean water provision (to ensure sufficient interviews were achieved with water only companies) and these were used as the sample frame - every customer therefore had an equal chance of being selected to take part in the research.

All water service users aged 18+ were eligible to take part in the survey providing they were aware of their water company either spontaneously or on prompting - respondents did not have to be the bill payer, although this information was captured in the survey.

Water companies were given the option to pay to boost their customer experience samples from 200 to 400 interviews. Most water companies chose to boost their number of interviews, the exceptions being:-

- Portsmouth Water and South East Water in Wave 1
- Portsmouth Water, South East Water, South Staffs Water and South West Water in Wave 2

Broad quotas were set on gender and age (three bands – 18-29, 30-59, 60+) to be reflective of each company's demographic profile. These were set based on 2011 Census data for the local authorities served by each water company.

We found in Wave 1 that the 18-29 age group and the younger end of the 30-59 year old group proved exceptionally difficult to find using the RDD approach, therefore we adopted a face-to-face top-up with younger customers to ensure we achieved sufficient interviews with 18-29s, and that the 30-59 age group wasn't too biased towards the upper end in terms of age. These were carried out in 2-3 locations within each water company area. Sample obtained from the electoral roll⁶ was also purchased as older age quotas began to fill to enable targeted sampling of younger customers.

In Wave 1, 1,249 face-to-face interviews were carried out at multiple locations in each water company area with an average of 73 younger customers per water company (numbers varied from 60 to 80 – an average of 19% of the sample per company). Combined with the 5,218 telephone interviews achieved, the total Wave 1 industry sample size for the Customer Experience survey was 6,467.

⁶ The electoral roll (or electoral register) lists the name and address of everyone who has registered to vote

In Wave 2, 1,186 face-to-face interviews were carried out at multiple locations in each water company area with an average of 70 younger customers per water company (numbers varied from 48 to 85 – an average of 20% per company). Combined with the 4,785 telephone interviews achieved, the total industry sample size for the Customer Experience survey was 5,971.

Due to some small discrepancies in the demographic balance of interviews achieved in each area, results were weighted to be representative of each water company area in terms of gender and age.

4.2 Recommendations For Shadow Year

4.2.1 Methodology

We recommend adopting a predominantly telephone-based survey using a Random Digit Dial (RDD) sampling approach to achieve a broadly representative sample, supplemented by targeted electoral roll sampling and face-to-face interviews to ensure sufficient interviews are achieved with younger age-groups.

There should be an equal number of face-to-face interviews carried out for each water company amongst 18-44 year olds, representing 20% of all interviews (i.e. 40 interviews, 30 with 18-29s and 10 with 30-44s. This will ensure fairness across water companies. Interviews should be carried out in 2 locations for each water company each quarter and interviewing locations should be rotated across the water company region.

In the pilot, if a respondent was not aware of their water company, they were prompted with the water company we believed they were supplied by, based on the water company flag attached by our sample provider. However, there were concerns within the industry regarding the accuracy of these flags. Our recommendation is therefore to ask respondents who do not know their provider for their postcode and identify their water provider using the Water UK postcode checker or a similar database to ensure we are prompting about their actual water company. If they are in any doubt the interview should not proceed.

Each water company should supply the appointed agent with full postcode information for all clean water customers on an annual basis. Water users will be approached using a random digit dialling approach, therefore they need to be able to name a water company to evaluate. If they don't know their water company spontaneously (and many users may not immediately if they are not the bill payer), they will need to be prompted, as mentioned above. In order to ensure sufficient interviews are achieved with water only companies, and to avoid confusion, in areas where different companies supply waste and clean water we recommend the water user is prompted about their clean water provider only.

As for the pilot, all water service users aged 18+ should be eligible to take part in the survey, providing they are aware of their water company either

spontaneously or on prompting; respondents do not have to be the bill payer.

Quotas should be set on gender and age (18-29, 30-59, 60+) to be reflective of each company's demographic profile based on 2011 Census data for the local authorities served by each water company.

Concerns have been raised by some of the water companies in the South of England that they have a higher proportion of more upmarket, ABC1 consumers and that these customers are less likely to award high scores for satisfaction; it has been suggested that quotas be applied on socio-economic status in line with the national demographic. Whilst an analysis of the Wave 2 pilot data for the entire water industry revealed that ABC1s are less likely than C2DEs to award a score of 9-10, there was no significant difference (at the 95% level) in the overall mean satisfaction scores awarded by each group. As such, and given that it will increase the difficulty of achieving the interviews and therefore the cost of the survey, we do not recommend imposing a further quota on socio-economic status.

4.2.2 Survey Frequency

As sampling is carried out independently of water companies, we recommend interviewing on the Customer Experience survey is carried out continuously over the quarter. This provides an incentive for companies to provide good customer service consistently and also mitigates against one-off events which otherwise might skew results. Results would be reported quarterly.

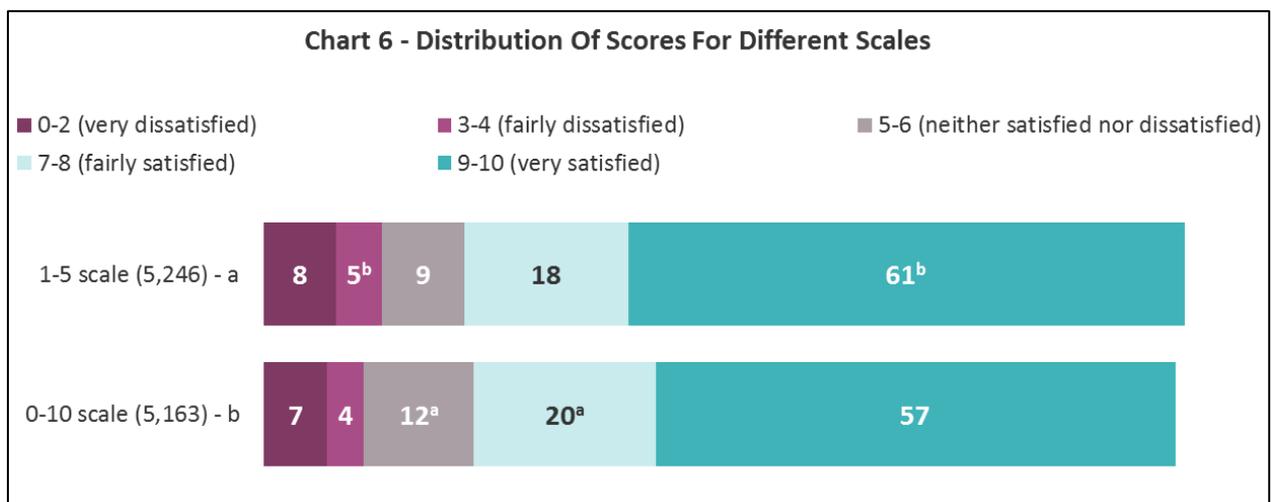
4.2.3 The Questionnaire

We recommend the Customer Experience Wave 2 pilot questionnaire is adopted for the shadow year.

5 CHOICE OF SCALE

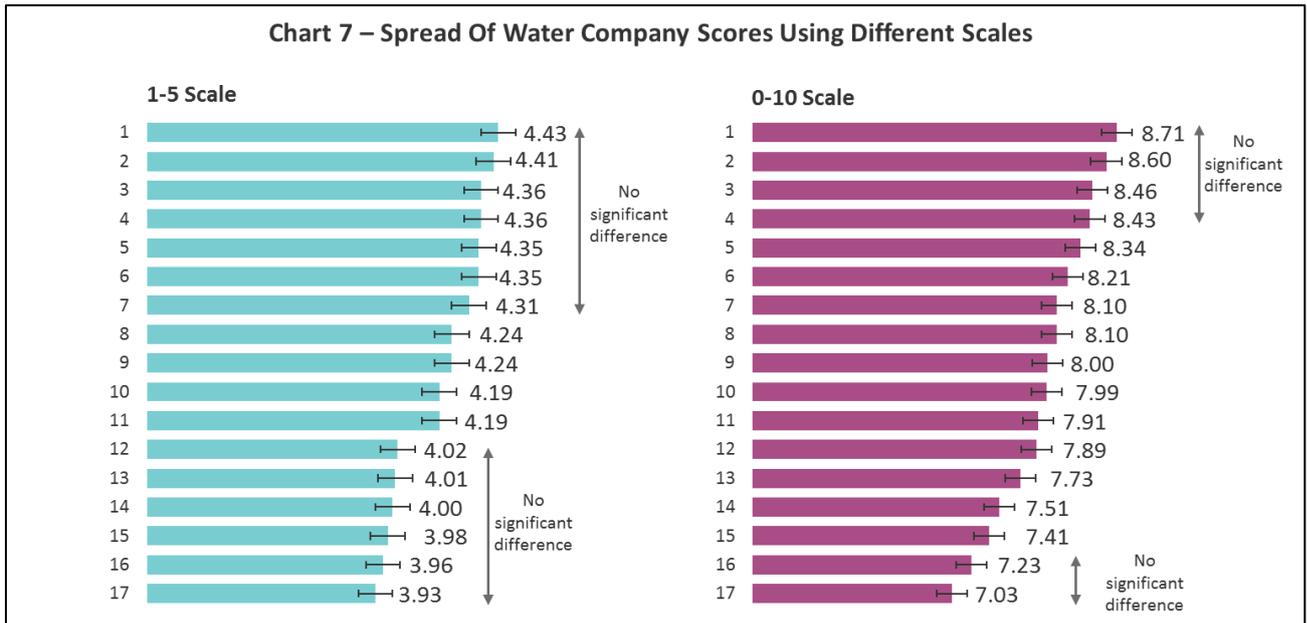
One of the issues with the SIM survey league tables is that scores for water companies have tended to converge over time leaving little to discriminate between companies in statistical terms. It was hypothesised that extending the range of scores customers can award for their satisfaction with their water company should provide greater discrimination between water companies. To test this, in the Wave 1 pilot, half of respondents were asked about satisfaction with their water company using a 5 point scale from very dissatisfied to very satisfied and half were asked on an 11 point scale from 0 (extremely dissatisfied) to 10 (extremely satisfied).

The distribution of results using a 1-5 labelled scale and a 0-10 scale is shown below.



On the industry level base sizes we are using to compare the scales, even a small difference is statistically significant, hence the differences at most levels of the scale despite results being broadly the same. It is unsurprising that those using a 1-5 scale gave a significantly higher proportion of very satisfied scores - their options are either fairly satisfied or very satisfied so they opt for 'very' as they feel they are more than 'fairly satisfied' with query handling, even if they are not absolutely delighted. However, on the 0-10 scale, the increased differentiation means that consumers have the ability to express their delight if necessary (with a score of 9-10) but can also give a score that indicates a high level of satisfaction that is not quite delighted e.g. a score of 8, hence the significantly higher proportion of respondents providing 'fairly satisfied' equivalent responses than for the 1-5 scale. This suggests that the 0-10 scale should be more discriminating between water companies and give more of a spread of results than seen in the SIM when using a 5 point scale.

The chart below shows the (anonymised) Customer Service league table of results amongst respondents using each scale.



It can immediately be seen that with a 5-point scale, water company scores fall into three broad groups - those achieving the highest scores, those with the lowest scores and those in between. There is no significant difference in results between any of the top 7 scoring companies or any of the bottom 6 scoring companies. With a 0-10 scale, whilst the top 4 and bottom 2 companies do not achieve significantly different scores there is a much greater spread of results, resulting in increased differentiation between companies.

Our recommendation is therefore to adopt the 0-10 scale for C-MeX.

6 SAMPLE SIZE AND COSTS

6.1 C-MeX Shadow Year Core Surveys

The issue of sample size is of key importance for C-MeX. With the move to in-period financial incentives, obtaining a robust annual sample size for each water company is essential to provide reassurance regarding the accuracy of results. Given the desire to offer customers a survey method in line with their contact channel on the Customer Service survey, and the fact that telephone remains the most common contact channel for all but one company, a robust number of telephone interviews is required for each company with customers using this channel.

However, customer contact through digital channels is growing and as online interviews are relatively cost-effective to administer, it is possible to substantially increase sample sizes for most water companies by emailing contactors with a link to the survey. That said, online contact is overwhelmingly billing related, so these surveys mainly act to increase the robustness of billing results rather than clean or waste water ones.

Over the two C-MeX pilot waves, on average water companies achieved 250 **additional** interviews per wave through online surveys, however there was a large variance in numbers across companies from an average of 25 interviews per wave to an average of around 400 per wave. This was due to the limited amount of digital contactors available (sometimes coupled with a lack of email address details) for some water companies.

The table below shows how sample size is likely to vary across the industry for various different telephone sample sizes per survey per water company per quarter, extrapolating the number of online interviews achieved in the two pilot waves to an annual figure (and assuming the smaller water companies supply two weeks of online contactors each month). It also specifies the Confidence Interval (degree of potential error) around results based on an observed percentage in the population of 50%. The final column gives an indication of the annual cost (excluding VAT) involved in adopting each sample size.

Telephone sample size per survey	Average sample size		Lowest sample size		Highest sample size		Cost £
	No.	CI +/-	No.	CI +/-	No.	CI +/-	
100	1,803	2.3%	998	3.1%	2,438	2.0%	£[<]
150	2,203	2.1%	1,398	2.6%	2,838	1.8%	£[<]
200	2,603	1.9%	1,798	2.3%	3,238	1.7%	£[<]
300	3,403	1.7%	2,598	1.9%	4,038	1.5%	£[<]

It can be seen that larger sample sizes have a lower Confidence Interval, which means that individual results will have less error around them, and are therefore a more accurate representation of the likely results within the population as a whole. Moreover, a lower Confidence Interval means that a lower difference in results is required between two companies (or for the same company from one wave to the next) to be able to say their scores are statistically significantly different. It is important to be able to differentiate between companies as far as possible in the league table – if Confidence Intervals are high, there will not be sufficient differentiation between companies in the table, and if rewards and penalties are based on league table position, some companies may receive a reward and others not despite having C-MeX scores that are not statistically significantly different.

The annual sample size for SIM is 800 interviews (Confidence Interval +/- 3.5%), with the SIM incentive based on performance over a four-year period ie 3,200 interviews (Confidence Interval +/- 1.7%). To achieve the same level of confidence in-period for C-MeX would require carrying out 300 telephone interviews per survey per company per quarter: this will inevitably be more costly. However, unlike in SIM, water companies will have the opportunity to achieve a performance payment or avoid an underperformance payment each year so there is greater potential for them to benefit from the incentive mechanism in at least one year of the price review period.

Our recommended sample size is 200 telephone interviews per survey, per company per quarter with 150 telephone interviews per survey per company per quarter offering a second option if supplemented by additional online interviews.

200 telephone interviews plus online interviews offer highly robust annual results for the majority of companies at substantially lower cost than for the 300 interview option (and with only a slightly higher confidence interval on average). It also allows for a reasonably robust sample size for each contact type (i.e. 67 Billing, 67 Clean, 66 Waste interviews). In addition, Ofwat's Final Methodology states that '*SIM will not operate in 2019-20. Instead we will run C-MeX [shadow year] that year. Some companies have reputational incentives for SIM in 2019-20. We will use the contact survey part of C-MeX to proxy the SIM...*'. As such, a sample size of 200 telephone interviews on the Customer Service survey, split 50% billing and 50% operations, will allow a direct comparison of results with the qualitative element of the SIM (please note that this will require converting results from a 0-10 scale to a 1-5 scale). Combining this with complaints data, to replicate the quantitative part of SIM, will enable companies and their stakeholders to assess whether they have met their reputational incentives for SIM in 2019-20.

Increasing the number of online surveys sent out to Customer Service contactors would be one way to address the reduced robustness of a 150 telephone interview sample, however the variation by water company in achievable numbers (as outlined in Section 3.2.1) prevents online interviewing substantively replacing telephone interviewing for C-MeX across the industry – on current volumes every water company should be

able to achieve enough online interviews to replace 50 telephone interviews per quarter but some would not be able to achieve many more than this. Moreover, the overwhelming majority of digital contact (89%) is billing-related, so online interviews could only really replace telephone interviews within billing. This suggests a quarterly sample size of 150 Customer Service telephone interviews if a minimum overall quarterly sample size of 200 Customer Service interviews is desired for all water companies.

If this approach is adopted, we recommend dividing the telephone interviews between billing, clean and waste water to give a minimum 100 interviews of each of billing and operations (once online interviews are included), as follows:-

- WASCs - 50 billing, 50 clean, 50 waste
- WOCs – 50 billing, 100 clean

Given that even for billing, telephone remains the most common channel for customers contacting their water company, and that telephone interviews are the means by which the views of these customers are assessed, it is important that a robust number of telephone interviews is achieved for each contact type.

In addition to the fact that some companies will not be able to achieve a minimum 200 Customer Service interviews per quarter, reducing the telephone interview sample size further, to 100 interviews, will severely limit the robustness of customer satisfaction results for all contact types. In this case, interviews would need to be split by contact type each quarter (to give a minimum 75 billing and 75 operations interviews per quarter, once online interviews are included) as follows:-

- WASCs - 25 billing, 38 clean, 37 waste
- WOCs – 25 billing, 75 clean

Clearly conducting only 25 interviews per quarter with telephone contactors for billing is not ideal, given telephone is still the majority channel of contact for billing, and is likely to introduce greater volatility into the results over time.

We note that reducing telephone interview sample sizes may reduce the comparability of Customer Service results with SIM results for those companies with reputational incentives for SIM in 2019-20. Online interviews could be used in their place with a correction factor applied (as mentioned in Section 3.2.3), but these results will be less comparable as there are no online interviews in SIM

It will not be possible to replace any Customer Experience surveys with online surveys as the only source of online contact details (other than going to the water companies themselves) would be to use a panel provider, so interviews would not be completely random and would be biased towards those who are more comfortable using technology and active online. Moreover, panel providers will not have sufficient panellists in many

WOC areas to be able to carry out sufficient interviews with different customers every quarter.

6.2 Validating The Correction Factor

As outlined in Section 3.2.3, we recommend a large-scale test during the shadow year, using a telephone methodology to achieve interviews with a sample of 1,000 digital contactors across all water companies and spread over time, to account for seasonality, and comparing this to digital contactor interviewers achieved online in order to verify the exact factor to use. The additional cost of this element will be £[~~2~~]plus VAT.

7 CALCULATING A C-MEX SCORE

This section deals with the subsidiary objective of exploring the weightings of the customer service and customer experience surveys within the C-MeX calculation and determining whether NPS should form part of the financial incentive.

In its final methodology for the 2019 price review, Ofwat stated that C-MeX would be retained as a 'single, but composite measure' of the Customer Service and Customer Experience surveys. One of the purposes of the pilot was to help Ofwat to determine what weightings should be applied to the components of C-MeX within the composite C-MeX score and also on which survey questions the financial incentive should be based (i.e. customer satisfaction only or customer satisfaction and NPS). Various options have been proposed following the pilot and these are outlined below:-

Option 1 - Base Case: composition based on 50%/50% Weighting of Customer Service (CS) & Customer Experience (CX) surveys using the results of the question on customer satisfaction (CSAT).

C-MeX score = (50% CustService CSAT) + (50% CustExp CSAT)

Option 2 - Base case adapted to have a slight majority weighting to one survey 60%/40%

- **Option 2a: 60% CS /40% CX:** composition based on a slightly higher weighting of Customer Service survey.

C-MeX score = (60% CustService CSAT) + (40% CustExp CSAT)

- **Option 2b: 40% CS /60% CX:** composition based on a slightly higher weighting of Customer Experience survey

C-MeX score = (40% CustService CSAT) + (60% CustExp CSAT)

Option 3 - Customer satisfaction with NPS element: composition based on the following weightings –

C-MeX score = (40% CustService CSAT) + (40% CustExp CSAT) + (20% combined CustService and CustExp NPS)

Option 4 - Base Case with Likelihood to Recommend (LTR): composition based on 50%/50% Weighting of Customer Service & Customer Experience surveys using, for each element, the average of the results of CSAT and LTR that is:

C-MeX score = (50 % CustService (50% CSAT + 50% LTR)) + (50% CustExp (50% CSAT + 50% LTR))

The league tables arising out of each of these options is indicated in the following tables based on combined results from Waves 1 and 2 of the C-MeX pilot.

As these are the results of a pilot, please note that they may not necessarily reflect company performance.

Option 1: 50% CS/50% CX			Option 2a: 60% CS/40% CX			Option 2b: 40% CS/60% CX			Option 3 - 40% CS/40% CX/20% NPS			Option 4 - 50% CS (CSAT + LTR) + 50% CX (CSAT + LTR)		
		C-MeX Rank			C-MeX Rank			C-MeX Rank			C-MeX Rank			C-MeX Rank
Industry		79.8	Industry		79.6	Industry		79.9	Industry		66.6	Industry		79.2
Portsmouth Water	84.9	1	Portsmouth Water	85.0	1	Portsmouth Water	84.6	1	Portsmouth Water	72.4	1	Portsmouth Water	85.1	1
Welsh Water	83.4	2	Anglian Water	83.2	2	Welsh Water	83.6	2	Welsh Water	71.4	2	Welsh Water	83.5	2
Northumbrian Water	83.3	3	Northumbrian Water	83.2	3	Northumbrian Water	83.5	3	Northumbrian Water	71.0	2	Northumbrian Water	83.5	2
Anglian Water	83.0	4	Welsh Water	83.1	3	Anglian Water	82.7	4	Anglian Water	70.3	4	Anglian Water	83.0	4
Wessex Water	82.8	5	Wessex Water	83.0	5	Wessex Water	82.5	5	Wessex Water	69.9	5	Wessex Water	82.8	5
South Staffs Water	81.5	6	Yorkshire Water	81.1	6	South Staffs Water	81.9	6	Yorkshire Water	68.8	6	South Staffs Water	81.1	6
Yorkshire Water	81.5	6	South Staffs Water	81.1	7	Yorkshire Water	81.8	7	South Staffs Water	68.7	7	Yorkshire Water	80.9	7
United Utilities	80.5	8	United Utilities	80.7	8	United Utilities	80.1	8	United Utilities	67.3	8	United Utilities	80.6	8
Severn Trent Water	79.8	9	South West Water	79.8	9	Severn Trent Water	80.1	8	Severn Trent Water	67.0	9	Bristol Water	79.4	9
Bristol Water	79.4	10	Severn Trent Water	79.5	10	Bristol Water	79.6	10	Bristol Water	66.2	10	Severn Trent Water	79.2	10
South West Water	79.3	11	Bristol Water	79.2	11	Hafren Dyfrdwy	79.1	11	Hafren Dyrdwy	65.5	11	South West Water	78.2	11
Hafren Dyfrdwy	78.4	12	Hafren Dyfrdwy	77.6	12	South West Water	78.8	12	South West Water	65.2	12	Hafren Dyfrdwy	76.5	12
SES Water	77.3	13	SES Water	76.5	13	SES Water	77.9	13	SES Water	63.3	13	SES Water	75.5	13
Affinity Water	77.0	14	Affinity Water	76.4	14	Affinity Water	77.4	14	Affinity Water	62.7	14	Affinity Water	75.0	14
South East Water	75.5	15	South East Water	75.3	15	South East Water	75.6	15	South East Water	61.2	15	South East Water	75.0	14
Southern Water	75.2	16	Southern Water	74.9	16	Southern Water	75.4	16	Southern Water	61.1	16	Southern Water	73.7	16
Thames Water	74.2	17	Thames Water	73.8	17	Thames Water	74.5	17	Thames Water	60.2	17	Thames Water	73.3	17

It can be seen from the various league tables that whilst company scores change under each C-MeX calculation option, the ranking of water companies is remarkably similar across each one. This would suggest that any of these options could legitimately be adopted.

There are various pros and cons to using NPS as a measure per se:-

NPS Pros

- As a standalone measure, NPS is simple to calculate and easy to understand
- It is a widely used measure so NPS results can be compared with those from other organisations and sectors (if publicly available). In this way it could be used as an alternative cross-sector comparison in C-MeX, given that the cross-sector threshold calculation is based only on the C-MeX scores of water companies with sufficient responses in UKCSI (12-13 companies in the 2018 surveys) and is of direct relevance only to those companies in, or close to, the top three.

NPS Cons

- NPS is a 'low resolution' measure as the 11 points on the scale are condensed to three in the final calculation (promoters, passives and detractors), resulting in reduced differentiation compared to Customer Satisfaction and potentially greater volatility making it more difficult for water companies to perform consistently.
- In total, 1 in 7 customers (16%) are not answering the question in the way it is intended. They either feel unable to answer the question (around 7%) or give a qualifying answer based on the fact that they can't currently change their water provider (9%). This, and other analysis we have carried out on NPS in a monopoly situation suggests customers with the latter attitude often opt for a lower score, which puts them into the detractor 'pot' and results in a lower NPS score compared to sectors where there is a true choice.

Moreover, there are some factors to consider in using NPS in the C-MeX calculation (Option 3). These are outlined below

Using NPS in The C-MeX Calculation: Pros

- Compared to not including it, the inclusion of NPS in the C-MeX calculation results in water company C-MeX scores that are marginally further away from the cross-sector threshold, resulting in a slightly more challenging target for water companies.

Using NPS in The C-MeX Calculation: Cons

- Increases the complexity of the C-MeX calculation and makes it more difficult for water companies to understand how overall C-MeX scores can be improved.

- NPS and Customer Satisfaction are, on the whole, measuring the same thing – based on the two waves of the pilot, NPS scores were highly correlated with customer satisfaction scores, with correlation factors of between 0.75 and 0.84 across the different surveys and waves in the C-MeX pilot (a correlation factor of 1 would mean the questions are measuring exactly the same thing, and of 0 that they are measuring completely different things). As such, NPS is adding very little additional differentiation to the C-MeX calculation.

The strategic principles of C-MeX call for the incentive to be simple and meaningful for companies and customers, whilst also stretching for water companies. Incorporating NPS into the calculation would certainly increase the stretch for companies, but the fact it is measuring a very similar thing to customer satisfaction brings into question whether the inclusion of NPS is enhancing C-MeX or making it unnecessarily complicated. On balance our recommendation is to adopt Option 1, a simple average of the customer satisfaction scores from the two surveys comprising C-MeX, whilst retaining NPS in the survey as a way of measuring cross-sector performance.

8 HIGHER PERFORMANCE PAYMENTS

8.1 Proposed Conditions For Higher Performance Payments

This section deals with the subsidiary objective of exploring the criteria required for water companies to achieve higher performance payments.

In the PR19 Final Methodology, Ofwat specify that in order for a water company to achieve the higher performance payments (between 6% and 12% of allowed residential retail revenues) it must meet all three of the following conditions:

1. The company performs **at or above a threshold based on cross-sector performance.**

This threshold could be the upper quartile all-sector performance on the UK customer satisfaction index (UKCSI) converted to a C-MeX equivalent, to ensure higher performance payments are only available for truly stretching levels of performance.

2. The company is demonstrating **satisfactory complaints performance**, as measured for example by the number of complaints received per household.

This approach will avoid the (perhaps unlikely) situation in which Ofwat awards a performance payment to a company that is achieving strong satisfaction scores on C-MeX by serving most of its customers very well, but which is serving a small proportion of customers poorly and receiving a high number of complaints per household, relative to other companies.

3. The company is **one of the top three companies**, by highest C-MeX score, **if more than three companies meet the first two criteria.**

Ofwat are capping the number of higher performance payments (between 6% and 12% of allowed residential retail revenues) to three per annum.

8.2 The Cross-Sector Comparison Gate

Focussing on the cross-sector performance threshold, Ofwat's proposal is to compare water companies' C-MeX scores to the **UKCSI all sector upper quartile (ASUQ)** score and award up to 3 companies higher performance payments if they surpass this threshold AND the satisfactory complaints threshold.

The intention is to set the cross-sector threshold each year, using the following scaling formula to convert the UKCSI upper quartile into a comparable C-MeX score. The following data is based on the results of Waves 1 and 2 of the C-MeX pilot and the UKCSI in 17/18.

$$\text{C-MeX ASUQ} = \text{C-MeX Mean} + (\text{UKCSI ASUQ} - \text{UKCSI Mean}) / \text{UKCSI SD} * \text{C-MeX SD}$$

Data Source	Calculation
C-MeX	Mean and SD will be calculated annually based on the scores of the previous year. To allow for comparison, only the scores of companies participating in the UKCSI will be used in the mean and SD calculation.
UKCSI All Sector UQ (population)	Taking the upper quartile of all company scores in the UKCSI questionnaire from the two past questionnaires (e.g. for 2018 it would be January 2018 and July 2017) and averaging the two.
UKCSI water companies (sample)	Mean and SD are calculated from the water companies participating in the UKCSI. In 2018, 14 companies participated (that is, for 14 companies the sample size of customers choosing to rate them was sufficient to be included in the UKCSI results).

$$\text{C-MeX ASUQ} = 79.8 + ((80.6 - 75.3) / 2.5) * 3.2 = 86.6$$

As mentioned previously, please note that as the results that follow are based on a pilot, they may not necessarily reflect company performance.

None of the top three ranked companies in the C-MeX pilot would have achieved the cross-sector threshold of 86.6 (based on a 50%/50% weighting of Customer Service and Customer Experience – Option 1). However, the top performing company is only 1.7 from the cross-sector threshold in this example.

Rank	Company	C-MeX score	Distance from cross-sector threshold
1	Portsmouth Water	84.9	1.7
2	Welsh Water	83.4	3.2
3	Northumbrian Water	83.3	3.3

The table below shows how C-MeX scores changed from Wave 1 to Wave 2 across the water industry.

	Change in C-MeX (W1 to W2)
Industry	
South West Water	4.25
Portsmouth Water	2.5
Northumbrian Water	1.65
South East Water	1.3
Affinity Water	0.9
Anglian Water	0.35
South Staffs Water	0.2
United Utilities	0.05
Wessex Water	-0.05
Southern Water	-0.35
Yorkshire Water	-0.75
Thames Water	-1.75
Welsh Water	-1.8
Bristol Water	-2.6
Severn Trent Water	-2.8
SES Water	-3.5
Hafren Dyfrdwy	-5.3

It can be seen that if, for example, Portsmouth Water were to make a similar improvement from Wave 2 to the first wave of the shadow year as they did from the C-MeX pilot Wave 1 to Wave 2, they would be in line for a higher performance payment, providing their complaints performance was satisfactory. However, they would have to at least maintain that level of service for the full year in order to realise this; other companies still have some way to go to reach the threshold, so it is even more stretching for them to achieve and sustain these satisfaction levels.

Our recommendation is that the proposed design is adopted as representing a stretching but achievable target.

Appendix 1 – Addressing Working Group Comments

Issue	Respondent	Comment	How Addressed
Online correction factor	[X]	We broadly support this recommendation; however, we would like some more information to understand the mix of online contacts and whether or not this could be driving satisfaction/dissatisfaction. For example, some companies have a high percentage of 'other online' contacts and we are unsure of what would be included in this – if these contacts were to set up an online account or something that was a similar sort of online transaction, then this may attract higher satisfaction than those online contacts that are generated as a result of service failure. We wonder if it would be beneficial to analyse the breakdown any further so that we can understand what is driving any differences.	See Section 3.2.3. Further investigation to be carried out during the shadow year.
	[X]	<p>[X] supports the recommendations for correction factors:</p> <ul style="list-style-type: none"> • Apply to online results in the shadow year • Complete further CATI survey for digital contactors to verify correction factor <p>[X] Request: Provide for an annual review to validate any permanent application of this correction factor.</p>	
	[X]	We support the proposal to apply a correction factor to online results in the shadow year to ensure companies are not advantaged or disadvantaged by having a different proportion of customers interacting online. To be comfortable that the right correction factor is applied there should be a CATI survey for digital contactors during shadow year to verify the correction factor across a robust sample of customers. The distribution and any difference between companies of both the existing sample and future surveys should be shared and reviewed. This should be a dynamic correction based on an annual survey to determine the correction factor.	
	[X]	We support the option to apply a correction factor to the online scores. However, as discussed at the workshop, we suggest further surveys are carried out in the shadow year to check the conversion rate is still applicable, ensuring a sufficiently robust sample size.	
Weighting by contact type (billing etc.)	[X]	Weighting of billing/clean/waste. As discussed at the workshop the largest percentage of contacts are related to billing. However, to ensure that operational contacts are receiving good services, 50% of the score is based on operational satisfaction survey results, so currently clean and waste each account for 25% of the overall score. The recommendation for clean to account for 33% and waste 17% goes against the principle mentioned earlier. Does this change in allocation give the perception that the service customers receive regarding waste contacts is less important? We suggest to ensure consistency, simplicity and to ensure that waste contacts are receiving good service the weightings should remain as currently (50%/25%/25%)	See Section 3.2.7. Final weighting option to be decided by Ofwat
Issue	Respondent	Comment	How Addressed

	<p>[X]</p>	<p>As a WOC we are aware of the typically higher scores in waste water which we naturally don't see. However, we support keeping the measure as simple as possible across industry. We welcome the weighting down of waste water – being a smaller component of “water” queries than clean water. However, we would ask that you reconsider the proposal to keep with 50:50 for billing:ops. Within working group discussions and the recommendation paper we didn't propose exactly representing our volume split of 90:10 but did recommend moving closer to the reality the volumes show. As mentioned previously, we propose 60:40 or 75:25 would be a more representative weighting. The rationale for this is, given the high proportion of billing contacts, we naturally all spend more time dealing with these customer journeys, in response to customer demand. It is also clear that the majority of customer interactions with all companies relate to billing – significantly more than half. And so, we feel to ensure the most robust measure, this should be reflected to represent customer demand and company effort. These are the contacts we have greatest opportunity to influence and deliver customer service and so it would naturally make sense to weight this up. We ask that this is reconsidered</p>	<p>See Section 3.2.7. Final weighting option to be decided by Ofwat</p>
	<p>[X]</p>	<p>Within the weightings, I can also confirm that [X] favours a 50:25:25 split between Billing, Water and Waste. We believe that the current split for SIM works well and that whilst we see fewest waste contacts, they also have the highest and most profound impact on our customers if they go wrong. We believe that an equal weighting between water and waste is appropriate and we think it makes sense to maintain this. All WASCs have worked hard to improve focus on waste water and we believe reducing its weighting may negatively impact this across industry. We support keeping the current weightings – especially when the modelling shared shows that changing it made no difference to rankings, and so it seems an unnecessary change.</p>	
	<p>[X]</p>	<p>[X] supports the recommendation to retain the 50% / 50% billing and operations split.</p> <p>[X] objects to the recommendation to change the weighting for clean/waste where it results in an under representation of waste water customers: A weighting of Water 33% and Waste 17% unacceptably decreases the new measure's balance in relation [X]'s operational dealings with a significant proportion of its customer base, who are provided Sewerage services.</p> <p>The methodology for the Customer Experience survey is already predicated on Water supply only data provision and effectively excludes by half the total number of survey opportunities for the company for two million [X] sewerage service customers.</p> <p>In combination with drawing the Experience sample from water customers only, the C-MeX incentive for a company like [X] to provide excellent customer service to our wastewater customers is seriously undermined.</p>	

		<p>【X】's perception of the logic for waste water weighting reduction is that it does not favourably compare to that applied for the weighting between Billing and Ops. There is no recommendation to weight at a 90/10 contact split based on contact volume for billing / ops because the latter ends up under-represented, but the proposal does exactly that on wastewater. 【X】 recognises the importance of making sure that there are strong incentives to deliver good service to all of our customers and this recommendation undermines that.</p> <p>【X】 Request: 【X】 respectfully calls for the weighting of clean/ waste within operations to be retained as an equal split of 25% Clean and 25% Waste.</p>	
	【X】	Proposals to look at this weighting were based on trying to more closely reflect the volumes of contacts within the survey methodology and weighting. Given the options being proposed do not (and cannot without significant change in approach) represent the number of customers who interact with water companies on different issues, in our view the existing SIM weighting should remain in place for C-MeX.	See Section 3.2.7. Final weighting option to be decided by Ofwat
	【X】	<p>I believe that changing the relative weightings attached to billing, water and wastewater services is undesirable. The principle behind incentive design is to drive the right company behaviour to the benefit of customers. The current 50/25/25 weightings have proven to be effective in incentivising all companies to improve customer service across all service areas.</p> <p>Whilst changing the relative weightings of water and wastewater services might serve to appease some companies' concerns it has now been shown that there would minimal or no change in relative company rankings. At the same time changing the weightings heavily towards water serves risks unbalancing incentives on WaSCs to improve wastewater services, which is unlikely to be in customer interests.</p>	
	【X】	We were very encouraged by the work Allto did on weightings, particularly where they highlighted that entirely natural quotas would mean the customer service survey would be too focussed on the billing service. To be 100% confident that the 33:17 weighting between supply and waste is correct, we would like to see this reviewed through the shadow year to check it is typical of companies' dataset.	
Sampling frequency	【X】	We believe changing to surveying customer service once per month as opposed to once per quarter will ensure more accurate scores, smoothing out any overly positive/negative anomalous weeks	See Section 3.2.5
Issue	Respondent	Comment	How Addressed
	【X】	Customer Service 【X】 supports the recommendations proposed for sampling frequency: Monthly surveys – random week each month	See Section 3.2.5 and 4.2.2. We do not

		<p>Interview fieldwork within two weeks Timetable - Monday 10 a.m. notifications and next day 5 p.m. data submission Results reported quarterly [✗] Request: For companies to receive their individual completed scores each month, together with indicative averages. Customer Experience [✗] support the recommendations proposed for sample frequency: Continuous interviewing over quarter; Results reported quarterly [✗] Request: For companies to receive their individual completed scores each month, together with indicative averages.</p>	<p>recommend supplying data on Customer Experience monthly as sample demographics vary by month depending on timing of face to face interviews</p>
	[✗]	<p>Recommendation to complete a survey on a random week each month. We support this option as it ensures consistency and fairness. It also mitigates against one off events, as you have stated. As you've stated the results will be reported quarterly and this seems a good option. As part of this it needs to be decided if the "Ofwat do not contact" lists will be provided following the completing of the monthly survey or quarterly. Also will the call recording and raw company data, for the challenge process, be send quarterly or weekly.</p>	<p>See Section 3.2.5</p>
Challenge process	[✗]	<p>We are supportive of keeping this as it is today. We would recommend more sharing from the chosen research partner of challenges received and upheld</p>	<p>See Section 3.2.9.</p>
	[✗]	<p>[✗] support the recommendations to maintain a reallocations and challenge process: Reallocations to be based on respondent reason. Challenges process provided this covers both reallocation and scoring error corrections - over 3 day window after preliminary release of data and call recordings, and final data release after one week. [✗] Request: Challenges to allow for scoring error corrections. To ensure consistency in the challenge process please provide definitions for any essential criteria, i.e. non-appointed activity and what circumstances constitute private network issues.</p>	
	[✗]	<p>We are supportive of the reallocation and challenge process outlined for the Customer Experience survey</p>	
	[✗]	<p>We support that there should be challenge process and that the 3-day window currently works and should remain unchanged. However, there seemed to be a variation in the number of challenges made by companies when this was discussed at the workshop. We suggest there needs to be very clear guidelines, so companies know what they can challenge to ensure consistency.</p>	
Issue	Respondent	Comment	How Addressed
Customer Experience	[✗]	<p>Validation steps within the Customer Experience survey are welcome as they will provide greater accuracy on individual company performance.</p>	<p>See Section 4.2.1</p>

validation of water company	[X]	<p>It is important to ensure that surveys are completed with consumers who are able to recall who provides their water service, and results are allocated to the correct provider of water services. Customers who are not [X] clean water customers were surveyed as part of the wave 2 customer experience survey. This seems to be a particular problem in areas where there are different providers of water and waste water services and a high volume of incorrect recordings have been found. However, we have also noted problems with customers who own holiday homes. The survey process and data allocation need to be reviewed to ensure the occasions where this happens are as limited as possible.</p> <p>The impact of this not being resolved at survey or data allocation stage could be significant. It is essential that this is fixed or a robust challenge process is put in place. An option to help resolve this could to be follow the methodology CCWater use in running their surveys in relation to confirming the customer is able to recall their water company when prompted, and not continuing the survey if not.</p>	
	[X]	<p>We were pleased that Allto were going to have another look at how data for the customer experience survey is gathered and ensure the questionnaire robustly filters out customers outside of our water supply area. We strongly believe if we don't get this right at point of survey there will be a significant number of challenges from the industry. We provided examples of out-of-supply area customers to Allto from the latest pilot and are happy to work with them to make sure we get this right.</p>	
Customer Experience face-to-face interviews	[X]	<p>[X] support the recommendations for face to face interviews: Non-weighted representative samples Refreshed combinations and rotations of locations for face to face interviews</p>	See Section 4.2.1
	[X]	<p>As shown in the survey results, there is an impact on results dependent on the survey method. Therefore it is important that a quota and methodology (or weighting if this is not possible) of face to face and telephone surveys for different age groups are set out to ensure this is fair across companies.</p>	
	[X]	<p>In the last two waves I have spotted that the geographic spread of face to face interviews across the [X] region has been limited. There appears to have been a heavy preference for customers that live in larger urban conurbations, generally in the south of the [X] region. Given the small number of surveys that take place in each wave a fully representative geographic spread across all companies' regions would be difficult, but I think trying to designing survey locations so that as wide an area as possible is visited over the course of a year would be worthwhile.</p>	
Issue	Respondent	Comment	How Addressed
	[X]	<p>Face to Face Surveys. We agree with your approach that there is no need to apply weighting for the face to face survey. However, this is based on the assumption that exactly the same number of face to face surveys will be completed for each company.</p>	See Section 4.2.1

Demographics	[X]	<p>It was discussed at the session that there has not been a significant impact noted across the industry in relation to impact of scores from different sociodemographic groups. We are continuing our analysis, but this is not in line with what we can see from the data relating to our customers from both wave 1 and wave 2. Other analysis we have completed shows a significant impact from socio demographic groups. We would like to have the anonymised industry dataset in order to be able to review against this hypothesis. We suggest this is looked at to avoid any unintentional bias.</p> <p>As we have suggested in previous responses weighting against sociodemographic group as well factors that have been weighted in the methodology (age, gender) should be weighted in line with national demographics, rather than for the demographics of each company individually to ensure against a regional bias.</p>	See Section 4.2.1. Provided data to [X] to show no significant difference in CSAT results between ABC1s/ C2DEs at industry level. We feel demographics should reflect water company populations
Sample size	[X]	<p>I recognise the important trade off that needs to be made between sample size and cost. If we move forward with an 'in year' incentive based on an 800+800 survey set it is important to recognise the change in statistical significance of the results as compared to the old 4yr SIM methodology. The slides presented at the last workshop make a good start in trying to do this.</p> <p>As the uncertainty range on results changes it becomes increasingly important that care is taken when setting incentive dead bands. It is important companies only receive rewards or penalties when C-MeX scores are significantly statistically different to the industry average.</p>	See Section 6
	[X]	<p>With regards to your recommendation of 200 interviews per survey each quarter we appreciate your comment that increasing the sample size to increase the error rate by 0.2% increases cost by 50% and there has to be a balance between cost and the benefit. As companies are having rewards and penalties based on the results it's important the results are robust. If the sample size remains as recommended maybe you can consider the potential error range when setting any reward or penalty. For example, if two companies score 8.40 (company A) and 8.42 (company B) the error range for these scores (based on 0.2%) is Company A – range from 8.38 to 8.42 and Company B – range from 8.40 to 8.44. Company B would receive a larger reward than Company A even though statistically company A's score might be higher (due to the range of error i.e. company A score could be 8.42 and Company B score could be 8.40).</p>	
Issue	Respondent	Comment	How Addressed
	[X]	This is the main area where we remain concerned by the recommendation. While of course we do understand the financial constraints surrounding the measure, in our view we have a special opportunity here to really develop a customer measure which is robust and fair and most importantly of all, ensures that as water companies we remain as focused as possible on	See Section 6

		<p>improving customer service further still. Further, given the significant rewards & penalties linked to performance, we want to be fully confident that these are applied in a fair environment. We therefore feel it essential to know there is strong confidence in the accuracy of the measure and by calculating reward & penalties each year, we're going to be basing these decisions on fewer customers than SIM penalty/reward is based on today.</p> <p>We have discussed at length in the past in this forum one of the shortfalls of SIM being that each survey wave is based just on 200 customers. While we welcome the increase in customer feedback from the addition of the digital contact we are concerned that the comparison with SIM (200 customers x 4 times a year, x 4 years) calculation for penalty or reward, is showing C-MeX being less statistically significant. With the proposed calculation of reward & penalty being significant in value, we feel the robustness of the measure needs to be at least as good as, but ideally better than today, in order to give legitimacy to the new measure.</p> <p>While we understand the comment made in the slides: "C-MeX with online represents a substantial improvement in accuracy over SIM on annual basis" we would question this when applied to rewards/penalties, which were over 4 years with SIM but now moving to 12 months.</p> <p>Further, there is a quote of "0.2% improvement will be 50% more cost" but we don't feel this is accurate. It may be a 0.2% point difference but the difference between 1.9% and 1.7% is actually an 11% change, so the difference in degrees of error between the two examples is 11% and not 0.2%; this therefore presents a far greater difference.</p> <p>If there remains a strong view the sample size should not be increased we would ask that the principles surrounding reward & penalty are reviewed to ensure no risk of companies being wrongly penalised or rewarded.</p>	
	[X]	<p>We are concerned that a sample size of 200 will be less accurate than under SIM over four years. The size of the financial rewards on C-Mex are significant and a company's position in the league table is a key factor in determining who accesses these. Therefore, a robust sample size with high confidence levels is critical and cost shouldn't be a barrier to achieving this. If a sample size of 200 is agreed, we believe the standard deviation used for performance payments should be adjusted to reflect the robustness issue.</p>	
Issue	Respondent	Comment	How Addressed
	[X]	<p>We appreciate the current proposal is more accurate on an annual basis than the SIM and the additional costs to improve accuracy are disproportionate. This is acceptable providing the confidence levels are reflected in the annual assessment for reward and penalty.</p>	See Section 6
C-MeX weightings (Customer	[X]	<p>We support 50:50 customer service:experience. We do have concerns about NPS being used given the lack of relevance in this sector, as long as there is no household competition. We have also seen this in responses to the survey question, with the majority of comments from</p>	See Section 7

Experience: Customer Service)		customers expressing confusion with how to answer and relevance. We can understand an interest in tracking but we would not support it forming part of the C-MeX score, linked to reward & penalty	
	[X]	<p>Our preference would be: Option 1 or Option 2a.</p> <p>We recognise the importance of ensuring all customers are represented when trying to measure the experience and satisfaction of customers with their water company and value the introduction of the experience survey. Through analysing the responses from the surveys we see a distinct difference in the feedback being received from the experience survey to the service survey. In that those who have not had a recent interaction often struggle to provide a robust response, often reverting back to experiences from a number of years ago. Which presents challenges when trying to understand the current performance of the company.</p> <p>We do not feel the use of NPS is the most appropriate measure to be used when measuring customer experience within the Water industry. We feel this is reflected by the percentage of customers who refuse to answer this question (approx. 10%) and of those that do answer (12%) make reference to the lack of comparison and water being a monopoly, items which we have little ability to influence.</p> <p>We would also wish to seek clarification as to why incorporating NPS would add value based on the above findings. As this measure is being merged with two others it does not enable the ability to benchmark outside of the water industry. If Ofwat wish to continue to capture this measure, we feel this should be done on a reputational basis only and separated from the C-MeX calculation.</p>	
Issue	Respondent	Comment	How Addressed
	[X]	<p>We are comfortable that Option 1 – Base Case (50:50 weighting of Customer Service (CS) and Customer Experience (CX) is the simplest and fairest route. We believe that keeping the metrics and measurement models clear and straightforward so all companies can be clear with customers on what we do.</p> <p>We understand this would mean that NPS is a standalone reputational measure and want to be clear that [X] would support this and think NPS is a valid and useful metric to use a part of a suite of measures used in understanding customer experience, loyalty and perception. It is our vision to be national leader and we believe being able to benchmark against other companies,</p>	See Section 7

		regardless of sector, is useful and will add value. We note that we are in the minority here, but would like it noted that we see value in this metric.	
	[X]	<p>[X] support the recommendation to retain C-MeX as a single, but composite measure: Weightings - In the absence of any particular compelling alternative rationale emerging from the pilot activities [X] supports a C-MeX scoring calculation that is based on equal weighting of the two component surveys.</p> <p>Survey questions determining financial incentive – As Net Promoter Score (NPS) and Likelihood to recommend (LTR) remain fundamentally non-representational of the present non-competitive household market, at this time, [X] supports Customer Satisfaction (CSAT) as the principal question.</p> <p>[X] requests: Based on the above, [X] prefers the application of Option 1; composition based on 50% / 50% weighting of Customer service (CS) and Customer Experience (CX) surveys, using the results of the question on customer satisfaction (CSAT).</p>	
	[X]	In line with the feedback from the working group we do not support the use of NPS to calculate C-MeX performance outputs at this time.	
	[X]	NPS is the metric that we will be using as a business metric and we support the use of it as part of the C-MeX metric. We therefore support Option 3 as set out in the slides – the combination of customer satisfaction with the NPS element	
	[X]	As outlined in the sub-group's recommendations paper, we do not see the benefit of using NPS when there is no competition in the household market. We don't believe options three and four for calculating C-Mex scores will work as they respectively show a perceived drop in customer service and fail the simplicity principle. Options 2a and 2b could work but as these options don't materially change the scores from the base case, we would recommend using a 50:50 weighting between the two elements.	
Issue	Respondent	Comment	How Addressed
	[X]	<p>Net Promoter Scores</p> <p>During initial discussions on the merits of using a Net Promoter Score as part of C-MeX, a number of potential issues were raised by a number of stakeholders, most notably:-</p> <ul style="list-style-type: none"> • Customers may not understand the questions being asked given they lack choice in the water company that serves their home • Many people lack experience of other water companies' service levels, making a direct comparison between water companies difficult <p>Based on my review of the latest results and customer comments from the recent pilot waves I think evidence is emerging that these concerns were well founded.</p>	See Section 7

		<p>The customer dropout rate from NPS questions is high. Many customers are giving neutral scores (5 or 6) to the question. Customers comments on the reason for their answers to the question reference issues such as “I have not experienced other companies”, “not sure who else could provide it”, and “don’t know other companies to compare it to”. These results indicate that the hypothetical nature of the current NPS question is not getting as many meaningful responses from customers as hoped for.</p> <p>I think we are getting to the point where the NPS element of the survey should either be reformulated so as to remove the hypothetical element to the question, or stopped.</p>	
	[X]	Option 1 is our preferred approach as we consider it fair that companies who are middle of the pack do not receive a reward or penalty. This will still drive companies to improve their performance as the over time the trend of water companies performance increase. If companies performance remains static they are likely to move down the league table and therefore into the penalty zone.	
Cross-Sector Comparison gate	[X]	We are happy with the UKSCI being a reference point for comparing cross-sector. We agree the enhanced level threshold is reviewed regularly to ensure that it is both sufficiently stretching but also obtainable	See Section 8.2
	[X]	We understand the metric. We believe simplicity across the board is best. We believe that the threshold is sufficiently challenging.	
	[X]	The Cross Sector Comparison calculation is understood and we believe this will provide a challenging target for the industry.	
	[X]	<p>[X] supports the recommendations to align the award of higher performance payments to companies that meet the following three conditions:</p> <ol style="list-style-type: none"> 1. At or above a threshold based on cross sector performance 2. Satisfactory complaints performance 3. One of top 3 companies, by highest C-MeX score, if more than 3 meet the first two criteria. 	
Issue	Respondent	Comment	How Addressed
	[X]	We agree with your approach of using the UKCSI all sector upper quartile score to calculate the cross-sector comparison gate. We support your approach of using UKCSI rather than NPS. We consider the cross sector threshold score is sufficiently challenging as a company will have to make significant improvements to increase their C-Mex score by 2.	See Section 8.2
Standard Performance Payments and Penalties	[X]	<p>We still have some concerns about using the standard deviations to calculate where penalties and rewards would be due. The system needs to ensure that there is an incentive for the whole industry to improve. Using the standard deviation model could lead to a scenario where all companies see a decline in performance, yet rewards would still be available.</p> <p>In terms of the different thresholds (the red, amber, green and blue zones), we would prefer a model where the blue zone is reduced. This is based on the rationale set out in our response to</p>	For Ofwat to address

		<p>your methodology last year where we outlined that companies should be disincentivised from being average or static, to enable continuous improvement: <i>“It is appropriate that an average company with average performance would expect to incur penalties rather than rewards on its ODI package. Companies will be incentivised to manage this risk by ensuring they deliver for customers, & this may help allay customers’ concerns.”</i> We consider that these points should be considered in the context of where rewards and penalties are due within C-MeX.</p>	
	[X]	<p>As discussed at the session there has been a significant narrowing in the range of performance over time with SIM and the wave 1 and 2 C-MeX results showing a relatively narrow range of performance. Survey scores naturally have a confidence range, so that the position of companies in the league table could, due to the bunching of performance, be as much due to the robustness of the survey and sample sizes as actual performance. As performance becomes more bunched there is more need for robust survey results, which has consequences for the sampling approach and sample sizes to ensure appropriate confidence in the scores. It is also important that the range in difference between scores, and confidence in those scores, is linked to how the penalty and rewards are calculated, so that companies are not over rewarded or penalized for survey characteristics. We therefore support an approach of using standard deviation rather than simple ranking to allocate rewards & penalties.</p> <p>In line with the comments above, our view is that the threshold needs to change from the 2SD used in AMP5 to a lower threshold so that the change in reward or penalty of a SIM point in AMP5 is broadly equivalent to the equivalent change in score for C-MeX in AMP7. From the options presented, Option 1 (assuming the blue zone is a neutral, i.e no reward or penalty, zone) is most consistent with the narrowing of the performance range and bunching of industry performance around the mean.</p>	
Issue	Respondent	Comment	How Addressed
	[X]	Without understanding the levels of payment or reward attached to each of the colour codings in your model, it is difficult for us to discern a preferred option. Is it possible to provide companies with more information on how this will be applied?	For Ofwat to address
	[X]	<p>[X] appreciate the merit of the alternative recommendations to determine which companies receive a performance payment or penalty.</p> <p>On balance [X] see the ranking option as simpler and easier to track, but could results in some rough justice with large differences in rewards/penalties for immaterial differences in scores.</p> <p>[X] Request: That standard performance payments and penalties are converted into revenue adjustments using Option 3; Thresholds based on groups of four (ranking), but with a check that differences are material and with companies given the benefit of doubt where marginal.</p>	

	[X]	We support the concept of outperformance and underperformance payments and so they should apply to as many companies as possible. Option 3 does this best and should mean that as many companies as possible are incentivised to push for better service to customers.	
Miscellaneous	[X]	Methodology: we are pleased with the multi-channel approach as a better representation of customers' reality.	N/A
	[X]	<p>Exclusions: [X] highlights two exclusion categories which require clarification to ensure that sample data submission is consistently provided by all companies - Non-appointed activities and private networks issues. [X] seeks definition and examples of non-appointed activities (see also challenge process request) and similarly for contacts regarding private network issues, to understand the degree to which such contacts are excludable from the sample provisions, to better determine the company classification and treatment of such contacts. The example questions below indicate the sort of practical uncertainties related to this area:</p> <ol style="list-style-type: none"> 1. If there is a 3rd party dispute between a tenant and landlord, in which the landlord neglected to fix an internal leak and has resulted in a higher bill issued. Should this be treated as a private 3rd party issue which [X] may exclude if they were unaware? 2. Contacts from non-account holders who identify themselves but don't live at the property and don't have permission on behalf of the customer. We will these be excluded from service survey data provision? <p>Excluding agency contacts under 1% rule - [X] also seeks confirmation of the approach to be taken in relation to the sample data provision for contacts handled by agents for the company. Under present arrangements [X] does not have the infrastructure in place to support the collation of agency data within the two-day timeframe following notification of a survey. [X] Request: That the exemption provision allowing agency contacts less than 1% be retained in the guidance for C-MeX.</p>	<p>We would recommend 1 should be included on the basis that the customer has had contact with the company about their bill, whether or not this is the water company's responsibility.</p> <p>2. Should be included if they were contacting on behalf of the customer. We recommend the agent compiles a list of any successful and unsuccessful challenges during the shadow year, such that these may be considered for inclusion in any company guidance for the period 2020-25.</p>
Issue	Respondent	Comment	How Addressed
	[X]	The only additional point is around the standard definitions for digital channels. We agree with this recommendation and suggest these should be included in the C-Mex guidance document.	Added to C-MeX guidance document.