Future Customer Service Incentive for Water

In association with Caroline Thompson Associates

November 2016

Economic
Insight

CONTENTS

1. Executive summary

This report for Ofwat was written in association with specialist consumer insight firm Caroline Thompson Associates. It considers how to provide water companies with incentives to improve their customer service levels. It specifically focuses on how a new incentive mechanism should approach measuring, evaluating and comparing companies' performance with respect to customer service outcomes.

The main findings of our work are:

(i) The main focus of the new incentive mechanism should be customer satisfaction. Asking customers directly about their satisfaction is the best way to measure this.

(ii) The scope of the survey should be extended to include customers that interact with companies through websites or self-serve.

(iii) The delay between customers' engagement with companies and being surveyed should be reduced. This could be achieved by using text and web-based surveys to enable instant feedback.

(iv) Cross-sector information could have a useful role in benchmarking firms' performance, but should be used in a considered way, alongside water sector-specific evidence.

1.1. Introduction

Ofwat introduced the SIM in 2010. There is widespread agreement it has strongly contributed to significant improvements in customer service in the water sector, including a 60% reduction in written complaints. Ofwat believes that there is scope for further improvements in firms' performance, and wants to ensure that it stretches companies to provide excellent customer service. By PR19, SIM will have been in place for two price control periods and, following feedback from companies and others, Ofwat committed in its May 2016 consultation to review SIM's effectiveness and consider whether it should be retained, modified, or replaced in PR19.

With its review in mind, Ofwat asked Economic Insight to *"identify and evaluate alternative approaches to measure, evaluate and compare company performance in terms of customer service, satisfaction, trust and confidence"*. This report sets out our answers to four overarching questions:

» What should be the objectives of a new incentive mechanism, and what outcomes should it help deliver?
» Given these objectives, when should an incentive mechanism be used?
» How should the outcomes of interest be measured?
» What are the options for using cross-sector evidence in the design and operation of the new incentive mechanism?

To help answer these questions, we have benefitted from the work of others, including:

» the practical and insightful input of Caroline Thompson and Helen Carruthers of Caroline Thompson Associates – they are qualitative and quantitative market research experts and have advised numerous companies (such as Diageo, British Gas and Costa) on the measurement of customer service performance and its drivers;
» the detailed research and analysis undertaken by companies and others; and
» constructive, challenging and open-minded input from the team at Ofwat, particularly during two half-day workshops.

In addition, we have completed various supporting research and analyses, including:

» reviewing the outcomes frameworks used by regulators in other sectors, including Ofgem;
» analysing the existing SIM data to help inform various questions relating to its statistical robustness and the implications of that for matters such as sample size; and
» interrogating third-party research and data of relevance to the measurement of customer service performance e.g. from the Institute of Customer Service's UK Customer Satisfaction Index (UKCSI).

There are, of course, some limitations in how far economic theory and the available evidence takes us in terms of answering the above questions. Where this is the case, we have sought to present our conclusions in a balanced way, applying some common-sense and judgement.

1.2. Findings and conclusions

1.2.1. Question 1: What should be the objectives of the new incentive mechanism and what outcomes should it help deliver?

Our view is that the primary objective for a new incentive mechanism should be to help mimic the incentives that firms face in "competitive markets" to deliver the outcomes consumers want from their suppliers in terms of customer service. In the absence of competition, and without such an incentive or alternative means of customer protection, firms may have an incentive to "under-deliver" service quality in order to reduce costs and increase profits. This would not be in the interests of consumers and could also undermine trust and confidence in the sector as a whole.

There are various ways in which SIM – in its current form – has helped to meet this objective.

» The first and most direct way is that it results in financial and reputational rewards to companies that have performed well and penalties to those that have not.
» The second way is that it encourages management to look and learn about how

other companies do things and make improvements accordingly.

While there have been significant improvements, in the absence of competition an incentive mechanism will continue to be necessary to encourage further improvements.

What outcomes should the new incentive mechanism help to deliver? The hierarchy below shows the connections between what companies do and companies deliver through to the high-level outcomes related to things that consumers feel.

Figure 1: Hierarchy of outcomes
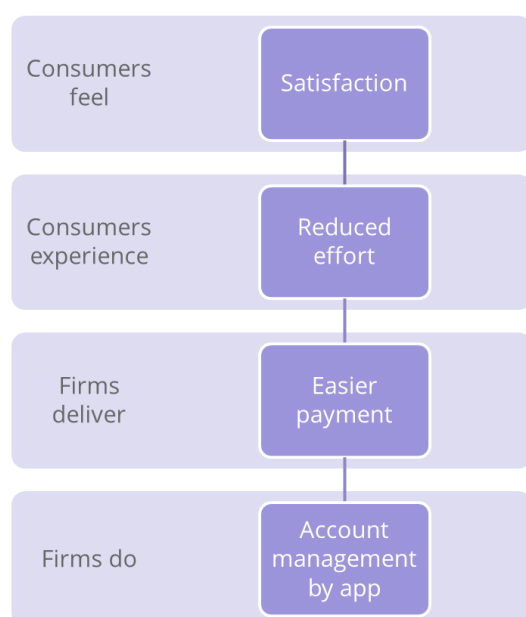
Source: Economic Insight analysis

Our conclusion is that the new incentive mechanism should focus on delivering the consumer-centric outcomes at the top of the hierarchy, such as increasing satisfaction, rather than the very specific 'actions' at the bottom of the hierarchy. This is for three reasons:

» first, those outcomes are likely to have the greatest intrinsic value to consumers;

» second, through their actions, companies have material control over them; and

» third, it leaves firms with greatest discretion over how they deliver the outcomes and, in doing so, it avoids the risk that the regulator codifies what "good customer service" looks like, potentially reducing the scope of

companies to meet the potentially differing needs of their respective customer bases.

The latter point is particularly important given the potential, set out above, for improvements and innovation in the ways that firms deliver customer service. If firms are to be encouraged to make greater use of, for instance, mobile technology, then this requires that they are able to experiment with new ways of delivering customer service.

We considered five specific customer-centric outcomes that firms in competitive markets target:

» customer satisfaction;

» customer trust;

» customer loyalty;

» customer perceptions of value for money; and

» fairness.

Our conclusions are as follows.

» The new incentive mechanism should incentivise excellent customer service by measuring customer satisfaction. The main reason for this is it is a broad outcome likely to capture the different aspects of service that customers care about. It will therefore leave room for companies to increase satisfaction by finding innovative new ways to increase their customer service quality.

» Trust and confidence are likely to be important to customers and there is a case for monitoring them or providing reputational incentives. Given their close connection to customer satisfaction, however, our view is that trust and confidence could be delivered by incentivising satisfaction alone.

» We do not think that Ofwat should incentivise customer loyalty. Our main reasons for this are: (a) arguably, customer loyalty is of greater relevance in the context of a competitive market where customers have choice and firms can invest in their brands; and (b) practically, measuring customer loyalty would be challenging without the opportunity for customers to switch. We do, however, recognise that some customers may derive some value from feeling "loyal" to a company even if they have no choice.

» The cases for focusing on value for money and fairness are not as strong as the case for satisfaction and, again, it is likely that they would be delivered by incentivising satisfaction alone.

Table 1: Summary of our conclusions on outcomes

Outcome	Assessment
Satisfaction	🟢
Trust and confidence	🟠
Loyalty	🔴
Value for money	🟠
Fairness	🟠

Source: Economic Insight analysis

1.2.2. Question 2: Given its objectives, when should an incentive mechanism be used?

Given the objectives set out above, there is a case for using an incentive mechanism when:

» there is no competition for a good or service (and possibly when competition for a good or service is sufficiently ineffective); and

» when such an incentive mechanism is the only or most effective tool for mimicking competition.

Accordingly, our views are as follows.

» If monopoly provision continues for household customers, the new incentive mechanism should apply to them from PR19.

» If retail competition is opened up for household customers, there may be a case for applying a separate incentive mechanism to wholesale businesses, depending on the materiality of their role in the delivery of customer service outcomes.

If service incentives are not applied in the event that retail competition is opened up, Ofwat could

nonetheless consider monitoring customer service performance (using its own or third-party measures of performance). It would be important to undertake this monitoring in a manner that does not distort the competitive process.

1.2.3. Question 3: How should the outcomes of interest be measured?

As noted above, our conclusion is that the new incentive mechanism should focus on customer satisfaction. There are various choices that it will have to make to do this, including:

» What should be the measure(s) or proxies of customer satisfaction?

» How can we ensure that the measures are statistically robust? In particular:

– Where surveys are necessary, which customers should be asked for their views?

– When should customers be asked for their views?

– How many customers should be asked for their views?

1.2.3.1. What should be the measure(s) or proxies of customer satisfaction?

At a high level we can distinguish between measures of consumers' views, such as their satisfaction, and measures of things that consumers do, such as making complaints. This corresponds to the difference between the 'qualitative' (surveyed satisfaction) and 'quantitative' (number of complaints and contacts) parts of the current SIM.[1]

With respect to "qualitative" measures, options include but are not limited to:

» asking customers to rate their service overall "Overall, how satisfied or dissatisfied are you with company X?" using a scale from very satisfied to very dissatisfied (numbered, for example, from 1-5, 1-6 or 1-10);

» asking questions to rate specific aspects of service "How satisfied or dissatisfied are you with company X's website?" (and potentially

---

[1] *The terms 'qualitative' and 'quantitative' are used differently by customer insight specialists. In the context of the SIM, 'quantitative' refers to the number of complaints and*

unwanted contacts, while 'qualitative' refers to the customer satisfaction survey.

derive an overall customer satisfaction score); and

» asking customers to offer their comments and analysing the text responses.

These options are not mutually exclusive and provide different types of information about customer satisfaction. The main distinction between the different measures is that some are for measuring performance levels, whereas others are for diagnosing it (i.e. establishing the drivers of performance).

Our research and the expert input of Caroline Thompson Associates suggests that asking customers to rate their service overall is the best way to drive improvements in company service. This provides all the information that is needed for regulatory benchmarking purposes. It avoids the risk that by asking questions on specific aspects of customer service, Ofwat implicitly prejudges what is and is not of importance to customers. It is broad enough to allow firms to innovate in their customer service provision.

There is a case for extending the scale against which customers rate their experience to include a sixth category of 'completely satisfied'. This could be useful in identifying the very highest quality service and addressing concerns about convergence in firms' scores.

We consider the pros and cons of other measures and measurement techniques, such as net promoter score, customer effort score and sentiment analysis, in the main body of this report. Our conclusion is that there is not a strong case for using these measures and measurement techniques for regulatory purposes.

We also consider the pros and cons of including 'quantitative' measures, such as the current measure of the number of complaints and unwanted contacts. We conclude that there is a case for removing the quantitative elements from the new incentive mechanism. The main reason for this is that it has the potential to give firms perverse incentives to make it more difficult for consumers to make contact.

1.2.3.2. Which customers should be asked for their views?

It is clearly important that *all customers* receive good customer service. The question here is a practical one, namely: who should we ask to get the information we need? Our conclusion is that the best approach is to extend the scope of the survey to customer contacts in its widest sense, including, for instance, self-serve and website users, without going as far as including the whole sample of customers.

This approach strikes the best balance between the benefits and costs of a wider sample. Asking customers who have had contact with their water company, rather than all bill payers, focuses on those that have experienced the customer service offered by their water company, and who are best placed to provide a view on customer service.

Extending the scope to those that have an interaction through websites or self-serve ensures that the whole range of customer experiences are covered, without the problems associated with samples of the entire customer base. Samples of the entire customer base will generally be less responsive to changes in firms' actual service levels, and so could reduce the incentive power of the incentive mechanism.

1.2.3.3. When and how should customers be asked for their views?

There are two issues here. First, how long after a customer has had contact with a company should he or she be surveyed? The answer to this question is "as soon as possible".

Too long a delay could result in: (a) customers simply forgetting what they thought; and/or (b) customers only remembering if they got particularly good or particularly bad services. Best practice in measuring customer satisfaction is therefore to bring measurement as close as possible to the point of experience.

Achieving instantaneous feedback from customers could be achieved by making innovative use of technology. For instance, rather than engaging in a telephone survey, customers could be asked for immediate feedback online or by text. This could also facilitate the broadening of the sample set out

above, to include website users. This would work better if the form of the survey was a short, single satisfaction question (and therefore excluded any detailed diagnostic information).

Second, how frequently should customers be surveyed? Our conclusion is that the current quarterly survey mitigates the risk of a single "poorly timed" survey that may unduly advantage or disadvantage specific companies, but that using technology as set out above could enable further improvements in frequency.

1.2.3.4. How many customers should be asked for their views?

Making innovative use of technology to change the way consumers give responses could facilitate material increases in the size of the sample. We also considered how adequate the current sample size was for the purposes of customer satisfaction measurement.

The current sample size for the qualitative SIM scores Ofwat publishes annually is 800 per company. To calculate the financial incentives at the end of PR14, Ofwat may use five years of data, resulting in a sample size of up to 4,000 per company for the qualitative SIM score (accounting for 75% of the overall score).

Some stakeholders have argued that larger samples are needed to improve the statistical robustness of the scores (particularly the annual scores) and so we undertook a statistical analysis of existing SIM data.

Our analysis shows that a sample size of 4,000 per company compares favourably with other well-regarded sources, such as the UKCSI and is sufficient to reliably distinguish between above, at and below average customer service performance. We therefore consider this sample size to be adequate for the purpose of setting financial incentives. This does not, however, preclude increasing the sample size if technology makes it easier to do so.

However, our analysis shows that, it is difficult to be confident that the measured variation in customer service performance between some companies is genuine or an artefact of statistical noise at the annual sample size of 800. As set out above, using technology more innovatively may allow the future incentive mechanism to

have a larger sample size than the current SIM, without a material increase in cost. Ofwat could also consider presenting additional information that would allow companies and their stakeholders to better understand the statistical robustness of annual scores e.g. presenting confidence intervals around the score.

Finally, we note that it is easier to statistically distinguish between the (say) high, medium and low performing groups of firms, than it is to distinguish between individual firms within the groups. For this reason, Ofwat could consider classifying companies into bands for the purpose of reputational incentives.

1.2.4. Question 4: What are the options for using cross-sector evidence in the design and operation of the new incentive mechanism?

Ofwat could decide to use cross-sector evidence to help determine the appropriate benchmark or threshold at which rewards and penalties would apply. Indeed, given that the objective of an incentive mechanism is to mimic the incentives that firms would face in a "competitive market", there is a good argument for looking at what firms in comparable competitive markets actually deliver.

Ofwat can choose between using cross-sector information in two broad ways.

» First, Ofwat could determine the threshold or benchmark only with reference to the customer service performance it observes in other sectors. For example, it could decide that water companies would only receive rewards if they achieve the customer service performance of the "top X" or "upper quartile" firms as measured by a third-party provider, such as UKCSI.

» Second, Ofwat could determine the threshold or benchmark with reference to a range of evidence, including but not exclusively the customer service performance it observes in other sectors. For example, it could set the threshold based on the average of the upper quartile performance of the water companies and the upper quartile performance of companies in other sectors. It could also use cross-sector evidence to inform whether (say) it would be appropriate to set the threshold or benchmark at the upper quartile of water

company performance or a higher (e.g. upper quintile), but without an "arithmetic" link to cross-sector performance.

Whether and how Ofwat should use cross-sector information to determine an appropriate benchmark or threshold, turns on four factors discussed in the main body of this report, namely:

» first, the extent to which the new incentive mechanism – without using cross-sector information – would provide incentives for firms to catch-up and shift the frontier of customer service;

» second, the extent to which other firms and sectors are comparable to the water sector;

» third, and relatedly, the risk of "unintended consequences" that could arise due to the variation in performance of individual companies or sectors; and

» fourth, the availability of robust and comparable metrics of customer service performance in other sectors.

Our conclusion is that, though cross-sectoral information could have a useful role in benchmarking firms' performance, there is and will be uncertainty over how comparable different sectors are to the water sector. Cross-sector evidence should, therefore, not be used exclusively by Ofwat, but rather help it inform a threshold or benchmark alongside other water sector-specific evidence (i.e. a variant of the second option set out above).

1.3. Next steps

Our analysis suggests that a new customer incentive mechanism could take the following form.

» A single question asking customers to rate their satisfaction, on a 1-6 scale.

» An increase in the scope of the survey to include all customers that have had an interaction, including website and self-serve users.

» Innovative use of new technology to allow customers to rate their satisfaction shortly after their experience, for instance by providing responses online or by SMS.

These changes would help the new incentive mechanism to meet future challenges and address the concerns and views expressed by stakeholders about the current SIM.

The potential new incentive mechanism is compared to the existing SIM in the table below.

1.4. Structure of this report

Our report is structured as follows:

» Chapter 2 presents background information on the SIM and the challenges it faces.

» Chapter 3 discusses the measures that the new incentive mechanism should focus on.

» Chapter 4 examines the best way to measure these outcomes.

» Chapter 5 assesses the role that benchmarking should play in the incentive system.

Table 2: Comparison of current SIM and new incentive mechanism

	Current SIM		Potential new incentive mechanism
Measure	Customer satisfaction (75%)	Contacts and complaints (25%)	Customer satisfaction
Survey channel	Telephone	-	Allow quicker responses, including online and text
Sample scope	Contacts	Contacts and complaints	Wider set of contacts and interactions, including self-serve
Sample size	800 per year	Full population	At least 800 per year
Sample delay	Up to four weeks	-	Up to one week
Sample frequency	Quarterly	-	At least quarterly
Incentives	Financial and reputational	Financial and reputational	Financial and reputational
Benchmark	Other water companies	Other water companies	Use information from other sectors

Source: Economic Insight analysis

2. Background and challenges

We consider the objectives, design and scope of the current SIM, and then discuss the reasons why a new incentive mechanism may be appropriate at PR19.

The purpose of an incentive mechanisms such as SIM is to mimic the incentives that firms face in competitive markets. It uses a combination of reputational and financial incentives, which are measured. Reasons to explore change include:

(i) The potential for further improvements in firms' customer service performance.

(ii) The possible introduction of retail competition.

(iii) Cross-sector developments in customer communications and customer service measurement.

(iv) Challenges that stakeholders have raised.

2.1. The current SIM

2.1.1. Objectives

The primary objective of the SIM is to help mimic the incentives that firms face in competitive markets, which encourage them to deliver the outcomes that customers want from their suppliers in terms of customer service.

It is one of a number of incentive mechanisms that Ofwat uses to encourage firms to improve service quality (which at PR14 were referred to as *'common performance commitments'*). Since its introduction, at PR09, there has been a marked improvement in company performance.

2.1.2. Design

The SIM uses a combination of reputational and financial incentives. Firms' customer service performance is measured every year, and their SIM scores are published. Firms will also get rewards or penalties based on their relative customer service performance across the period up to PR19. These rewards and penalties can range from reductions in household retail revenue of 12% to increases in household retail revenue of 6%.

Companies are given a score out of 100 for their performance, which is based on two separate measures. The qualitative measure, which accounts for 75% of the SIM score, is based on surveyed satisfaction among customers that have had direct interactions with companies. Consumers rate their satisfaction on a scale of 1 (very dissatisfied) to 5 (very satisfied).

The survey is managed by Ofwat, and is based on a sample of 800 customers per firm per year, spread across quarterly 'waves'. Half of the score comes from billing contacts, with the other half coming from operational contacts. For water only companies (WoCs) these are all water contacts, while for water and sewerage companies (WaSCs) this is split evenly between water and wastewater contacts.

The quantitative measure, accounting for 25% of the SIM score, is based on the number of complaints and unwanted contacts that firms

receive, weighted for their seriousness. Unwanted telephone contacts are given a weighting of one, with written complaints, escalated written complaints and complaints investigated by the Consumer Council for Water weighted by factors of 5, 100 and 1,000 respectively.

2.1.3. Scope

The current SIM applies only to household customers, and reflects their experience across the retail and wholesale parts of the value chain.

2.2. Reasons to explore change

2.2.1. Scope for innovation and further improvements in customer service

There is potential for firms to innovate and thereby improve their customer service performance. This could include greater use of mobile technology, for example with respect to how customers pay for their bills and manage their accounts. For example, Ofwat found that only two water companies allow their customers to manage their accounts through an app.[2] Evidence suggests that a significant proportion of consumers, and a majority of 16-24 year olds, prefer to use an app to a browser for paying bills online (and banking).

Figure 2: Preference for using an app versus a browser for banking and paying bills online
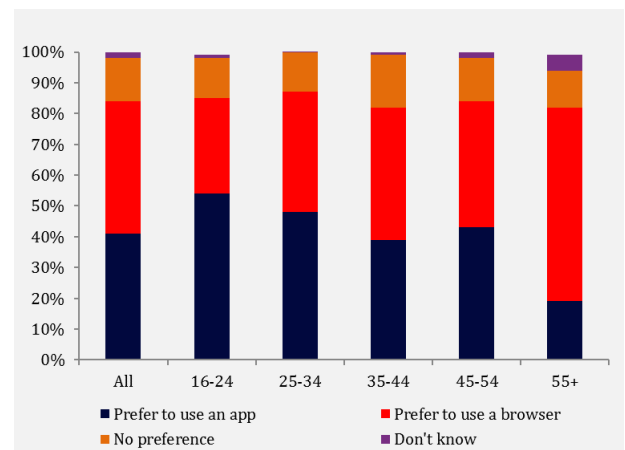
Source: Ofcom Adults' media use and attitudes report, 2016

[2] *'Costs and benefits of introducing competition to residential customers in England.'* Ofwat (2016);

Looking at third-party measures of water companies' customer satisfaction indicates that there is material variation between companies and across sectors in terms of reported customer satisfaction, as shown in the figure below. These could be caused by various factors, including differences in the marginal costs and benefits of service 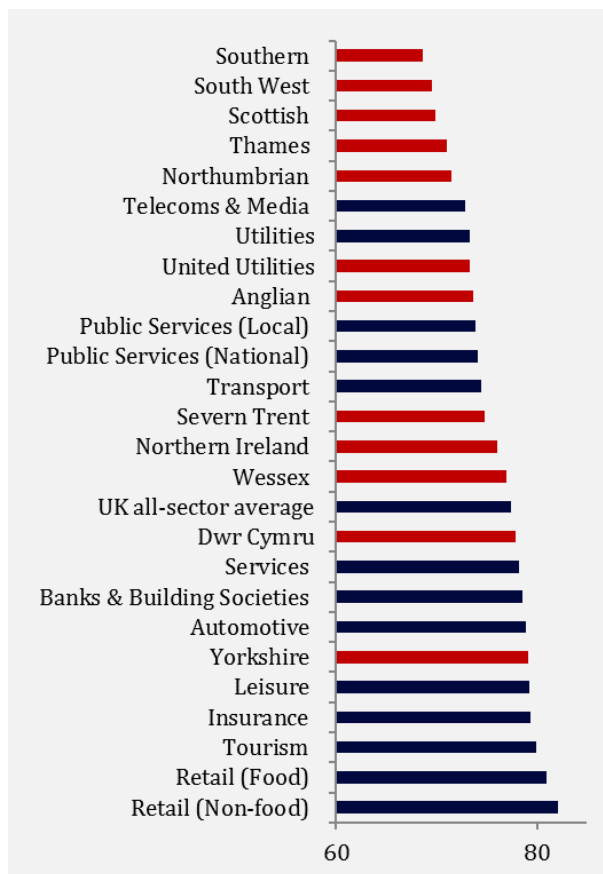improvement, but may also reflect scope for frontier improvements. We also note that some company scores are based on small sample sizes (a minimum of 35 per company).

Figure 3: UKCSI satisfaction scores for water companies and UK sector averages, July 2016

Source: Economic Insight analysis of Customer Satisfaction Index – Utilities Sector Results (The Institute of Customer Service, July 2016).

2.2.2. Possible extension of retail competition

Non-household customers will benefit from the opening up of retail competition from 2017. Government is now considering whether and, if so, when, to extend retail competition to

households. Ofwat estimates that this could generate net benefits ranging from -£1.4 to £2.9 billion over 30 years.[3]

The potential introduction of retail competition emphasises the importance of the scope of the new incentive mechanism and of understanding how it should apply across the value chain.

At present the SIM reflects the experiences of household customers. The customer problems that are measured within it could be due to issues in the retail part of the value chain (such as billing) or in the wholesale part (such as leaks). In their interaction with consumers, customer service could be 'provided' by the retail part (in the case of a retail problem), or by both retail and wholesale parts (in the case of a wholesale issue). The incentives themselves are applied to firms' retail revenues. This is shown in the diagram below.

Figure 4: How the SIM currently applies across the value chain
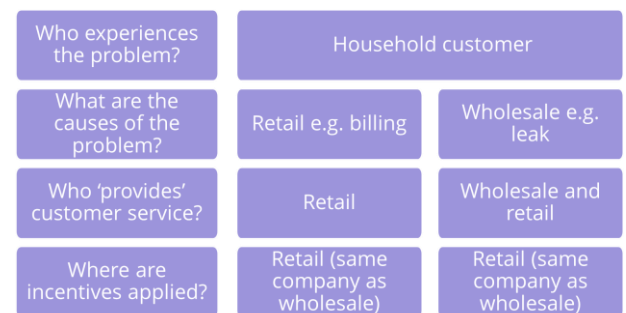
Source: Economic Insight analysis

If retail competition is extended to household customers, and if it is effective, retail businesses will have incentives to provide the outcomes that consumers value with the need to provide additional incentives through the SIM. We note that the SIM does not now apply to non-household customers, on the grounds that retail competition gives firms incentives to deliver the outcomes that customer value.

To the extent that wholesale businesses play a material role in the delivery of customer service

---

[3] *'Costs and benefits of introducing competition to residential customers in England.'* *Ofwat (2016);*

outcomes, it may be appropriate to apply a separate incentive mechanism to them.

2.2.3. Cross-sector developments

The challenges that the current SIM faces also need to be considered in the context of wider developments in consumer behaviour and the measurement of customer service.

For example, customers are increasingly using social media to interact with companies. Some companies are therefore concerned that the quantitative SIM score is at risk of being outdated with respect to the communication channels it includes, and potentially misleading due to excluding the views of some consumers. Others, however, worry that this could result in companies withdrawing or discouraging the use of social media as a mode of communication for fear that easier methods of communication would result in greater volumes of customer contact. [4]

Since the SIM was first developed, there also have been a number of developments in the measurement of customer service. In particular, customer effort score has emerged as a relevant measure of customer satisfaction, while firms are increasingly making use of social media to analyse their customers' feelings. We have consulted consumer insight experts to ensure that we take new developments fully into account.

2.2.4. Challenges raised by stakeholders

While the SIM has been successful in encouraging companies to increase their service levels, stakeholders have raised a number of potential challenges that a future incentive mechanism may face. These mostly relate to the best way to measure performance, and we address these issues in this report.

» Some stakeholders have queried the **scope of the sample**, suggesting that it is inappropriate to restrict the qualitative SIM sample only to those customers that have had direct contact with their company.

» **Convergence** in SIM scores has given rise to concern that small differences in scores will result in significant differences in penalties and rewards.

» Relatedly, it has been argued differences in firms' qualitative SIM scores could be due to **statistical noise**, suggesting that it may be necessary to increase the sample size.

» There has been disagreement over the **balance of qualitative and quantitative measures** in the SIM, with some arguing that the quantitative SIM should be dropped in order to increase focus on customer experience, and others arguing that the quantitative measure is necessary to give firms an incentive to prevent problems occurring in the first place.

In addition, there has been concern that **demographic and regional variations** give rise to differences in SIM scores that are outside of company control. This issue is outside the scope of this report.

2.3. Focus of this report

This report focuses on three main issues:

» Which outcomes should the new mechanism incentivise?

» How should the outcomes of interest be measured?

» How should companies' performance be benchmarked, and what use should be made of cross-sector evidence in doing this?

Additionally, we considered the appropriate objectives and scope of the new incentive mechanism. Our conclusions on these issues are as follows.

» The main objective of the new incentive mechanism is to help mimic the incentives that firms face in competitive markets.

» In relation to the appropriate scope of the new incentive mechanism, overall we think that:

– it is appropriate not to apply an incentive mechanism to non-household customers

---

[4] *'The future of the service incentive mechanism: a report for Affinity Water'. Frontier Economics (2016).*

(in England) because they are expected to benefit from retail competition; and

– it is appropriate to apply it to protect household customers, unless the Government decides to extend retail competition to them in time for PR19.

» Finally, in relation to whether the new incentive mechanism should apply to the wholesale, retail or both parts of the supply chain, our analysis suggests the following.

If household retail competition is not introduced, the new incentive mechanism should apply to the whole supply chain

– If competition is extended to household retail, there is a good case for removing service incentives from retail businesses. There may be a case for retaining an incentive mechanism for wholesale businesses, depending on the extent to which their actions remain necessary for providing the customer service outcomes that customers value.

3. Outcomes

Given our analysis of the appropriate purpose and scope of the new incentive mechanism, we have considered which outcomes are appropriate to focus on. This involved determining the properties that outcomes would have to satisfy, and assessing candidate outcomes against these properties.

Our analysis suggests:

(i)   Outcomes need to be intrinsically valued and within company control.

(ii) There is a strong case for incentivising customer satisfaction.

(iii) Trust and confidence are important, but may be better suited to reputational incentives or monitoring and may be delivered by incentivising satisfaction in any case.

3.1. Identifying outcomes

The first step in designing any incentive system for outcomes is to understand which outcomes it is appropriate to incentivise. To do this, we create a shortlist of candidate options and then evaluate them against criteria that we have developed.

3.1.1. Narrowing down options

We first developed a shortlist of potentially relevant outcomes by examining the following sources.

» Outcomes monitored by firms in competitive markets.
» Third-party measures, for instance from industry bodies.
» Water company outcomes and related performance commitments at PR14.
» Outcomes relating to customer service used in other outcomes frameworks.

This suggested a range of potential options, including satisfaction, effort, trust, confidence, loyalty, value for money, fairness, volumes of complaints, fairness, and quality and speed of service.

We then considered how **outcomes** relate to other things that firms do in the delivery of customer service. These occur across a number of levels.

» At the lowest level we can think of things that firms do. Although they might not be outcomes, per se, they are important determinants of outcomes. An example would be a firm giving its customers the ability to manage their accounts by app.
» Above this, we can think of things that firms deliver. An example would include easier ways to pay, the delivery of which depends on firms doing things such as account management by app.
» At a higher level still, outcomes could be thought about as comprising the things that consumers experience. For instance, firms delivering easier ways to pay would result in consumers experiencing less effort when they engage with their firms.
» Above this level, outcomes could be thought of as being things that consumers feel. This

includes their sentiments that they feel in response to the things that they experience, such as satisfaction in response to a reduction in the effort that they have to exert when dealing with a particular company.

These different levels are shown in the figure below. We note that there is often uncertainty as to where in the hierarchy any individual outcome will fit. For instance, value for money could be thought of as being at the level of things that consumers feel. It could also be thought of as something that customers experience, and which determines how satisfied they then feel.

Figure 5: Hierarchy of outcome levels
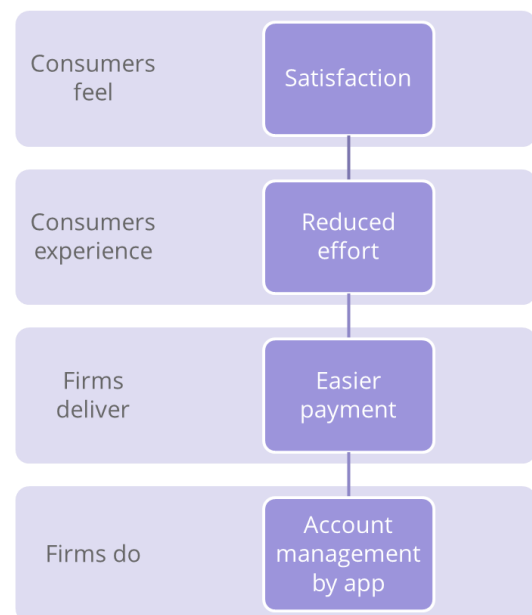
Source: Economic Insight analysis

We think that it is appropriate to focus on outcomes at the highest level in this hierarchy. These outcomes are likely to have the most value to customers, and incentivising them leaves firms with the greatest discretion over how they deliver good customer service, avoiding the risk that the regulator effectively 'codifies' what good customer service looks like, thereby reducing scope for innovation.

Our analysis of potentially relevant outcomes suggested that the following could be of relevance for the new incentive mechanism.

» Satisfaction
» Trust and confidence
» Loyalty

» Value for money

» Fairness.

3.2. Evaluation criteria

Having narrowed down a list of potentially relevant outcomes, we now consider the properties that outcomes subject to incentives should have.

3.2.1. Value

An outcome will only be appropriate for incentivising if consumers place **value** on it. This criterion is based on the observation (set out above) that under monopoly supply firms lack incentives to provide things that consumers value, to the extent that they incur costs in providing them.

Although we set out that costs, alongside value, will also determine the *extent* to which particular outcomes are underprovided, we do not think it appropriate to include a cost-related evaluation criterion. This is because costs determine the magnitude of the incentives that it is appropriate to attach to particular incentives, rather than whether it is appropriate to attach an incentive of any level.

In considering whether consumers place value on particular outcomes, we will need to consider the extent to which value depends on particular **contexts** or **market situations**. For example, firms in competitive markets often make significant investments in brand value and reputation. The value of this for customers is likely to be mostly associated with reduced search costs. Knowing that a firm has a reputation for the quality of its goods means that consumers do not have to spend time researching and shopping around. In other words, its value is *instrumental*, rather than intrinsic.

The observation that value may be context-dependent highlights the possibility that value differs across the different levels in the 'hierarchy' that we set out above. In particular, we think that, if an outcome is expressed at the level of things that consumers feel then it is most likely to accord with what they value, whereas

this is less likely for things that firms do, which are context-dependent.

Customers may value different things in the context of **wholesale services** and **retail services**. Alternatively, it may be that the **issues** that customers have differ across wholesale and retail, for instance a leak versus a billing problem, but value the same overall outcome, such as satisfaction. We will consider how value differs across these contexts when we assess potential outcomes.

We note that accordance with what customers value is particularly important in a context in which there is potential for innovation in what firms deliver. As set out in the previous chapter, there may be material scope for firms to improve their customer service by using new technological developments. Measuring outcomes at a high level will ensure that innovations that consumers value are captured by the new incentive mechanism.

3.2.2. Control

To be appropriate for incentivising, an outcome will also need to be **within company control**. This is important because attaching incentives to outcomes that are outside company control could result in perverse incentives, leading to consumer detriment.

If a regulator based incentives on that outcome that was outside company control, the firm would receive penalties and rewards that were unrelated to its actions. There would therefore be no link between the incentive mechanism and what the firm did, resulting in there being no incentive effect. For example, firms would receive rewards when they had done nothing to cause better outcomes, which would be paid for by consumers through higher bills.

In contrast to value, the extent to which outcomes are within company control is likely to **increase** the more one goes **down** the hierarchy of outcomes. This means that, for some outcomes, there may be trade-offs between value and controllability across the different levels.

This is because firms have direct control over the things that they do, whereas moving up one step to things that they deliver can introduce

some uncertainties associated with their ability to deliver them (for example a company might introduce account management by app, but fail to test it adequately with customers and thereby fail to deliver easier payment in practice). Moving further up could introduce additional potential variations that are outside company control, including external events that have the potential to affect how consumers feel.

This is not to say that firms do not have a meaningful level of control over, say, satisfaction; rather, that they have less control over satisfaction itself than the things that they do that contribute to satisfaction. There may, of course, be variation between firms in terms of the extent to which they choose to exercise control, which itself depends on their ability and incentive to pull the levers available to them. The extent of control can also vary over time, for example due to changes in customer service technology.

It is important that outcomes are **both** valued by customers and under company control. There are particular problems associated with placing incentives on an outcome that is under company control but not valued. In this case, firms would have incentives to focus on outcomes that were not valued by consumers, but consumers would still have to fund rewards for firms.

3.2.3. Coherence with the overall regulatory framework

3.2.3.1. Aligns with Ofwat's vision, strategy and Water 2020 principles

Ofwat's shared vision involves customers and wider society having trust and confidence in water and wastewater services. This is based on the understanding that customers and society want good quality service at a fair price, services where and when they want it, prices that are affordable and good value, and services and prices that are sustainable over the long term.

Ofwat's position is informed by first-principles economics and aligns broadly with those that we have set out. Overall, we think that the process we have set out to identify outcomes will provide outcomes that align with Ofwat's vision, strategy and Water 2020 principles. Ofwat's vision places particular importance on trust and

confidence, and we will consider these explicitly as potential outcomes.

3.2.3.2. Other performance commitments

Ofwat's evaluation criteria include a requirement that the new incentive mechanism adds value over other common performance commitments. While the common performance commitments from PR19 are yet to be determined, they are unlikely to focus directly on customer services, and so it is very likely that any customer service incentive will 'add value'. It may be that there is the potential for a customer service incentive to 'overlap' with any common performance commitments. In practice, this is very difficult to assess, but we have analysed data on the overlap of the current SIM with the existing common performance commitments. Further detail of this analysis can be found in section 2 of the annex.

3.3. Satisfaction

Customer satisfaction is the focus of the current SIM, and is one of the most widely used measures of customer service performance. It forms the basis of many other regulators' incentive mechanisms (for example, as part of Ofgem's Broad Measure of Customer Satisfaction) and monitoring efforts (for instance Ofcom's monitoring of communications markets). It is widely used by companies in competitive markets, and a number of third-party measures, such as the UKCSI, are available.

Satisfaction is expressed at the highest level of the hierarchy of outcomes, things that consumers feel. It is generally thought of as representing the extent to which firms meet the service expectations of their customers.

Figure 6: Example outcomes hierarchy for satisfaction
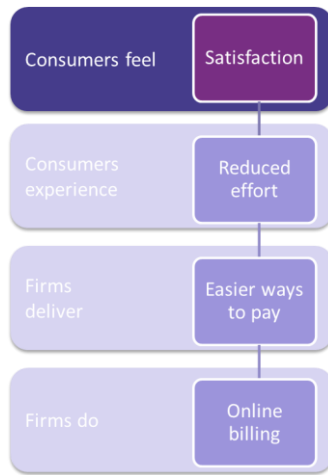
Source: Economic Insight analysis

By construction, we think that satisfaction is broad enough that it corresponds very closely with what consumers value, irrespective as to whether they are household or non-household and irrespective as to whether the service in question relates to firms' wholesale or retail activities.

For instance, suppose that household consumers place most value on low effort interactions with companies, whereas business customers place most value on experiencing high service quality. In contrast to outcomes expressed below this level, satisfaction **is** valued in both contexts.

This observation is also true when we consider different parts of the value chain. For instance, suppose that consumers value quick interactions in their billing contacts, but want dedicated service when they have water or wastewater problems. Again, satisfaction accords with value in both cases, in a way that outcomes expressed below this level do not.

Satisfaction also appears to show material correlation with trust and loyalty (as measured through the net promoter score). This is shown in the chart below. For water companies, the correlation between UKCSI satisfaction scores and trust was 93%, while the correlation between UKCSI satisfaction and net promoter score was 88%. To the extent that customers care about these other outcomes in addition to satisfaction, this suggests that attaching incentives to satisfaction may, nevertheless, encourage firms to provide these outcomes as well.

Again, we note that some company-specific scores had small sample sizes. However, this is not true of the sector scores, which also showed the correlations that we have illustrated here with trust and net promoter score.

The correlation with effort (12%) is positive, in contrast to what one would expect (i.e. one would expect lower effort to be associated with higher satisfaction). This could reflect that effort sits lower in the 'outcomes hierarchy', although it may reflect the influence of a small number of 'outliers'. Sector-wide effort scores, which have larger sample sizes, show the expected correlation.

Figure 7: Correlation between UKCSI customer satisfaction measures and trust, water companies and UK sector averages
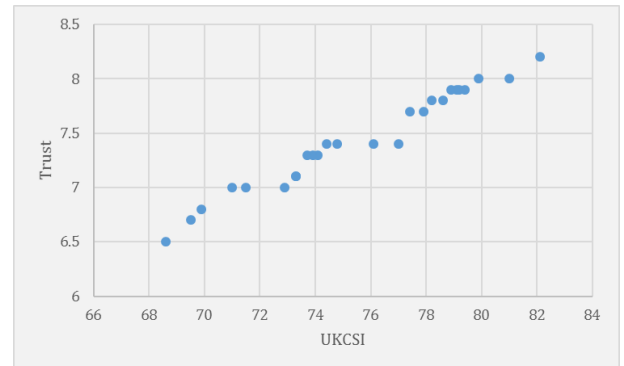
Source: Economic Insight analysis of UK Customer Satisfaction Index – Utilities Sector Results (The Institute of Customer Service, July 2016)

Figure 8: Correlation between UKCSI customer satisfaction measures and net promoter score, water companies and UK sector averages
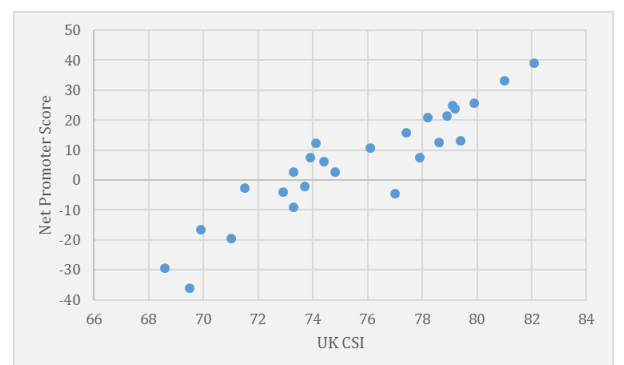
Source: Economic Insight analysis of UK Customer Satisfaction Index – Utilities Sector Results (The Institute of Customer Service, July 2016)

Figure 9: Correlation between UKCSI customer satisfaction measures and customer effort, water companies and UK sector averages
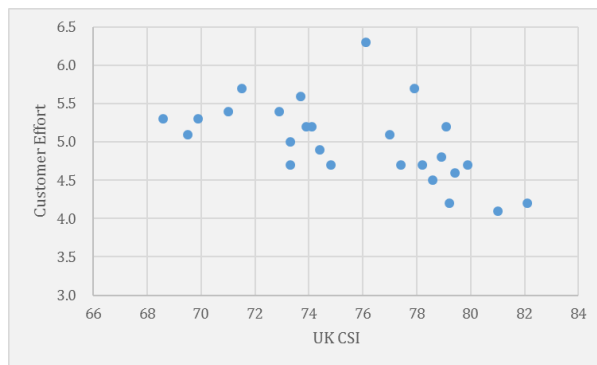
Source: Economic Insight analysis of Customer Satisfaction Index – Utilities Sector Results (The Institute of Customer Service, July 2016)

We also expect satisfaction to be strongly correlated with consumers' feelings about the value for money that they get. This is both because perceptions of value for money will be affected by satisfaction relative to what the consumer paid, and because price is likely to affect consumers' expectations of the service levels that they will receive.

Firms are also likely to have a control over satisfaction, even though it is expressed at a high level in the hierarchy. It will be determined by things that they deliver and are within their control. We note that satisfaction does not score as highly as outcomes that are further down the hierarchy, as things that consumers feel can be influenced by things that are outside consumer control. For instance, the extent to which a consumer feels satisfied may be affected by factors such as their general mood.

To assess the extent of control over satisfaction, we looked at evidence from UKCSI data as to the extent to which different sectors' satisfaction scores move together. If this happens, this suggests that factors outside of company control have induced these changes. This shows evidence of some common movements, for instance many sectors' scores fell in 2014.

Figure 10: Sector UKCSI scores, 2012-2016
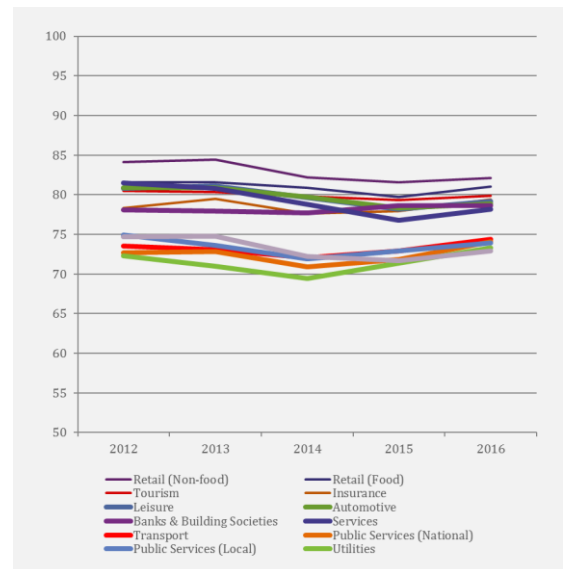
Source: Economic Insight analysis of UK Customer Satisfaction Index – The state of customer satisfaction in the UK (The Institute of Customer Service, July 2016)

To focus on this issue more closely, we indexed scores to 100 in 2012 and looked at changes in satisfaction for a sub-set of firms. This suggests that satisfaction is outside of company control, to some extent. While the common movement in 2014 appears indicative of external events influencing consumers' assessments of their satisfaction, there are no obvious candidates as to what such external events could be.

Figure 11: UKCSI scores, indexed to 2012 for a sub-set of sectors
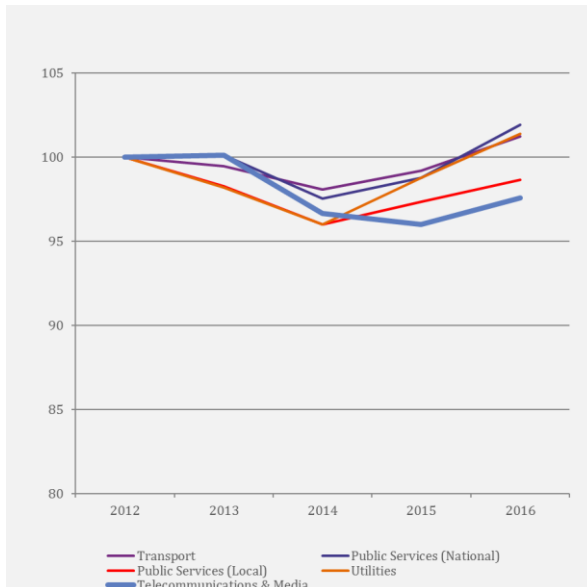

Source: Economic Insight analysis of UK Customer Satisfaction Index – The state of customer satisfaction in the UK (The Institute of Customer Service, July 2016)

The issue of the impact of external events draws attention to the possibility that there may be advantages to relative measures, which could control for changes in satisfaction that are the result of common external trends. We discuss this issue in more detail in chapter 5.

Overall, our assessment is that customer satisfaction scores very highly with respect to its accordance with what customers value, and highly with respect to the extent to which it is controllable.

Practical example: Whitbread – Premier Inn/Costa Coffee

Premier Inn and Costa Coffee operate in industries in which consumers have a wide range of choices. They seek to differentiate their brands through the customer service they offer, and measure this using their 'Listen & Learn' research programme. This uses a wide range of measures, which serve a range of purposes from performance monitoring, staff incentives and problem diagnosis. A further point to note is how using email surveys has enabled quicker responses than via a telephone survey.

Focus

Customer experience is monitored in both companies using two measurements:

» brand standards
» customer satisfaction.

The brand audit is conducted by a professional audit company at each location twice a year, and measures hard operating standards.

The indicators in the survey include:

» net promoter score and open text reasons for this rating;
» overall satisfaction with the visit (5-point scale);
» likelihood of revisiting (5-point scale);
» value for money (5-point scale).

They also include more detailed ratings across the customer journey. For example, for Premier Inn this includes the ease and speed of check in, the comfort of the beds, and the cleanliness of the rooms. For Costa Coffee it includes time spent waiting to be served, the choice of food available, the quality of the drink, and the tidiness of the coffee shop.

The measurement approach was updated in 2008, moving from a twice-yearly postal survey to a larger and more immediate survey in which a sample of customers are emailed a link within 24 hours of using the hotel or café. A portal is updated daily with the previous day's results.

Rationale

The survey is run for the following reasons.

» To keep staff focused on delivering excellent customer service;
» To diagnose service areas requiring attention. If the customer service measures drop, or fail to increase according to targets, then the other survey questions are used to shed light on reasons for this at either a company-wide level, regionally, or at a category level.
» To provide a measure linked to bonus payments. Targets are set for net promoter

score each year, which must be met to trigger bonus payments.

» To assess customers' reactions to new initiatives and products.

» To demonstrate to investors that the companies provide excellent service levels, which investors recognise has a positive impact on future use.

Ultimately, focus on customer service reflects the likelihood of consumers reusing the service or recommending the brand, and therefore enhancing future sales.

How the information is used

The key measures of net promoter score and overall satisfaction form part of the 'Customer Heartbeat', along with measures of sales, staff turnover, and other indicators that are regularly reported from board level down to team meetings. Annual targets are set and progress towards these is reviewed frequently, keeping the focus on service as something that each employee can contribute towards

Each site receives monthly feedback from their own customers. Targets are set on service measures and trigger bonus payments.

In addition to these 'top level' measures, the data is extensively used throughout the companies as a diagnostic tool. This includes feedback from 'free flow' commentary given after the net promoter score, which is analysed via sentiment analysis (new in 2016, and not yet operating across Premier Inn's survey).

Source: Caroline Thompson Associates research

3.4. Trust and confidence

Many companies in competitive markets monitor trust and confidence in their businesses. A number of third-party measures of trust are also available, including from Which? and the Institute of Customer Service. In a regulatory context, the Water Industry Commission for Scotland (WICS) also measures "esteem" in Scottish Water. There is no obvious way of distinguishing between 'trust' and 'confidence' at the level of things that consumers feel, so we use the terms interchangeably.

In its broadest sense, trust reflect a sentiment that is held across all consumers, whether or not they interact with the company in question. This is because it reflects both consumers' direct interactions with companies, and their expectations of the service that they would receive, in a variety of different circumstances.

In a narrower sense, trust could be framed with respect to particular interactions, in the same way that questions about customer service can be asked about a particular experience in dealing with a company. In this sense, trust is a feeling that is engendered by companies' performance with respect to identifiable interactions.

Like customer satisfaction, trust and confidence are expressed at the level of things that consumers feel. Again, this has the advantage that as an outcome it remains the same, even when the things that deliver trust vary across different consumers and different parts of the value chain, and evolve over time.

Figure 12: Example outcomes hierarchy for trust and confidence
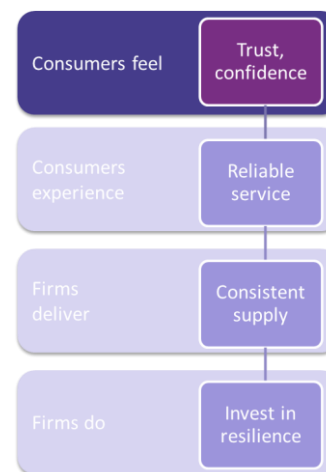
Source: Economic Insight analysis

We expect that trust and confidence are strongly valued by water consumers. Being unable to trust that they will receive a safe and reliable supply would have a material adverse effect on consumers' lives. Unlike satisfaction, however, we think that the value of trust and confidence may vary across different parts of the value chain. For example, consumers may place more value in having trust and confidence in firms'

wholesale functions (such as water safety) than retail functions such as billing.

We also expect that the value that customers place in having trust and confidence in their water services is discontinuous in its nature, particularly for wholesale services. For instance, suppose that a company was able to reduce the likelihood that its water was unsafe from 50% to 20%. Despite there being a material increase in water safety, it is unlikely that consumers would place much value on this while there was still a significant chance that the water was unsafe. Once this chance was eliminated, however, it is likely that value would increase markedly. Beyond this point, however, it is likely that the value of increases in trust would be modest.

Figure 13: The discontinuous nature of the value that customers place in trust in water companies
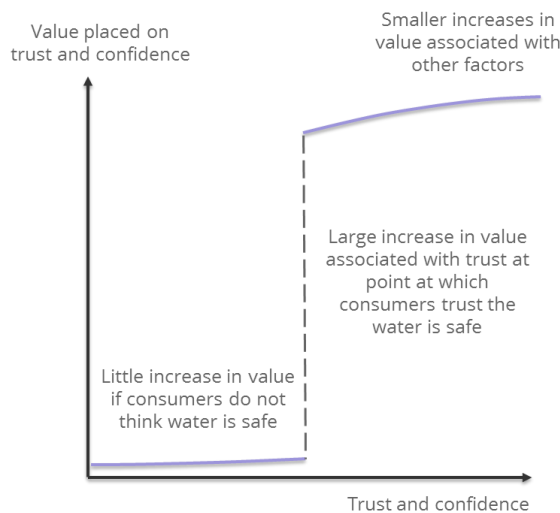
Source: Economic Insight analysis

On the other hand, in the context of retail activities, in which the focus is on activities such as billing, it may be that consumers do value marginal increases in the extent to which they have trust and confidence in companies.

We expect that there may be a material extent to which trust and confidence are outside of company control. This is because satisfaction is expressed with respect to companies' actual performance, whereas trust and confidence (partially) relate to consumers' expectations of how companies' would perform in certain circumstances. We note that there is some example that trust in some sectors move in step

with each other, in a way that they would not if changes were entirely within company control.

Figure 14: Industry-wide trust in Water, gas and electricity and food/grocery sector
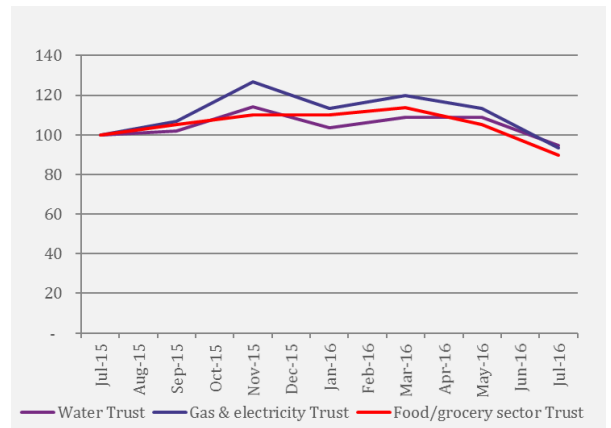
Source: Economic Insight analysis of Which? data

Overall, we think that trust and confidence are likely to have strong accordance with consumer value, though it may be that the value that consumers place on additional increases in trust and confidence in a wholesale context are more limited. The extent of correlation of trust and confidence with satisfaction suggests that providing incentives to companies to increase satisfaction would also have the result of increasing trust and confidence.

We think that this suggests that it may not be appropriate to attach **explicit financial incentives** to trust and confidence. Reputational incentives could be provided, so long as published trust scores were accompanied with appropriate caveats with respect to controllability.

Alternatively, Ofwat could monitor trust and confidence and publish figures at the level of the whole industry. This would avoid the problems associated with attaching separate financial incentives to trust and confidence, while allowing Ofwat to understand whether there were specific issues of concern.

3.5. Loyalty

In competitive markets, companies frequently measure customer loyalty. We have not, however, found any evidence of loyalty being measured by regulators, with the exception of

circumstances in which it is viewed negatively in the context of low switching rates. In practice loyalty is often measured using the **net promoter score**, which a number of third parties also monitor, though this measure is sometimes also used as a proxy for satisfaction.

Loyalty may be thought of as being expressed at the level of things that consumers feel, as shown in the diagram below. This could have meaning even in a monopoly context, if considered in the sense of whether the customer would choose a particular firm, if they had a choice, or if they would recommend a firm. Alternatively, however, we could think of loyalty as being at a different level, that of things that consumers do. This does not have meaning in the context of monopoly supply.

Figure 15: Example outcomes hierarchy for loyalty
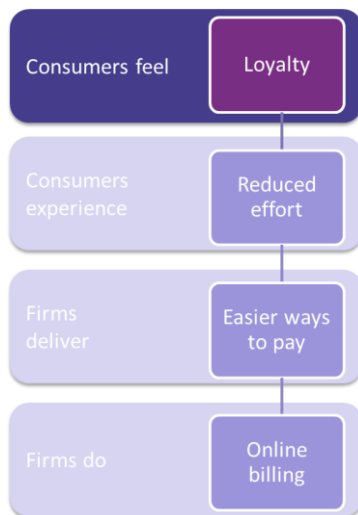
Source: Economic Insight analysis

Although it is expressed at a high level, we do not expect that loyalty is of material value to consumers, though it is likely to be correlated with things that they do value. For instance, by delivering things that make them feel trust or satisfaction, firms are likely to engender feelings of loyalty in their customers, even if this feeling of loyalty may not be valuable in itself.

With respect to control, we expect that firms are likely to have a material amount of control of the loyalty that customers feel, precisely by delivering things that they do value. Overall, therefore, we think that loyalty accords poorly with value, though it is within company control to some extent.

Practical example: Metro Bank

Metro Bank has had a strong emphasis on customer service, and in 2016, Metro Bank won five awards at the Moneywise customer service awards. It emphasises turning customers into "FANS" who recommend Metro Bank to others.

Again, we note that the measures are used for a range of different purposes. In this example, technology is also used to ensure that outcomes are measured 'in real time'.

Focus

Metro Bank runs a multi-channel 'Voice of the Customer' programme, capturing customer feedback from all channels: in store, telephone, social media, online and app. Net promoter score is the main measure used, along with associated reasons for customers' scores. It is measured every day for every store.

Rationale

The measure is used for a number of reasons:

» To keep staff focused on delivering excellent service.
» To be make sure the bank is recommended, so that positive word of mouth brings more customers
» To demonstrate the above to investors.
» To judge branch performance, and focus management effort where it is most needed.

How the information is used

The measure is used both to judge and reward branch performance, and is used as a diagnostic tool to identify any issues quickly and place appropriate resources where needed. A customer satisfaction-related bonus is awarded to employees annually.

*Source: Caroline Thompson Associates research*

3.6. Value for money

A number of firms included value for money as performance commitments at PR14, and it is monitored by the Consumer Council for Water as

a part of a range of indicators. We also note its inclusion in a number of other outcomes frameworks that we looked at, though it is often used as though it were synonymous with 'efficiency'. Perceptions of value or money are also monitored by firms in competitive markets.

At a high level value for money represents consumers' feelings about the service level that they received related to the amount that they paid for it. We note that this is high-level and, like satisfaction, likely to be broadly applicable across the value chain and across different customer types.

Figure 16: Example outcomes hierarchy for value for money
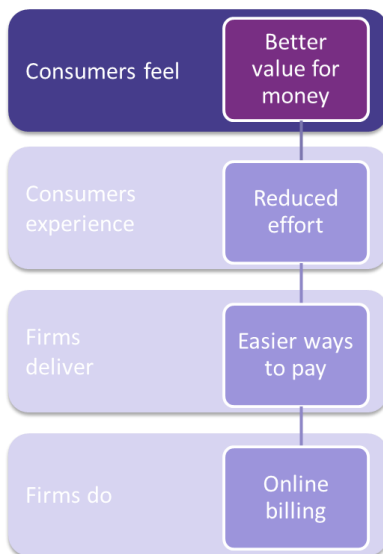
Source: Economic Insight analysis

Importantly, there are various means within the price control process to ensure that value for money is delivered i.e. that there is the right balance between price and service. Arguably, the cost assessment process is designed to ensure that the "price" part of the value for money equation is "right". Therefore, our view is that the new incentive mechanism should "add value" by addressing the other part of the equation: the service level. It is of course important that the two parts of the price control are designed in a consistent way.

Figure 17: Customer satisfaction and value for money in water, 2008-2015
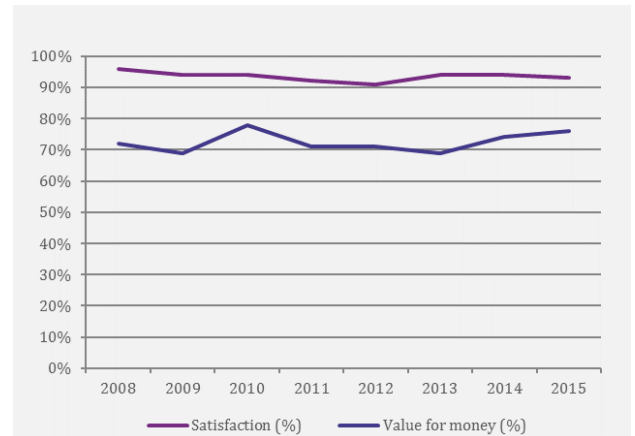
Source: Water Matters: House Customers' Views on their Water and Sewerage Services (Consumer Council for Water)

3.7. Fairness

Fairness is central to a number of regulators' approaches, and a number of companies' performance commitments could be thought of as relating to fairness. For example, Ofgem has a licence condition which states that licensees would not be treating customers fairly if their actions or omissions 'significantly favoured' the licensee's interests, and were likely to give rise to detriment to the customer.[5] A number of water companies have specific performance commitments that broadly reflect considerations of fairness.

The Financial Conduct Authority (and, previously, the Financial Services Authority) has rules relating to the fair treatment of customers. These relate to: consumers having confidence that fair treatment is at the heart of firms' cultures; products and services being targeted at consumers they have been designed for; and the provision of clear information. They also stress consumers receiving suitable advice, products performing as consumers have having been led to expect, and an absence of barriers to switching or complaining.

For a regulated monopoly, with prices that cannot vary arbitrarily between customers, unfair behaviour is most likely to occur with respect to the quality of service provision. For

---

[5] *'Implementation of the domestic Standards of Conduct – decision to make licence modifications.'* Ofgem (2013).

example, it would be unfair for consumers with visual disabilities only to provide account management online.

Seen in this way, being "fair" or "unfair" towards customers is related to the availability and quality of what firms do and deliver for different types of customers. It is lower down the value hierarchy compared to, say, satisfaction. That is, it is likely that "fair" or "unfair" behaviour by firms would give rise to some customers expending greater effort than others to deal with their supplier and, therefore, experience lower levels of satisfaction.

Accordingly, for the reasons set out above, it would be more appropriate to incentivise satisfaction than fairness. This leaves it open for companies to choose to increase their customer satisfaction by engaging in doing and delivering things that customers view as being "fairer".

3.8. Conclusions

Overall we think that there is a clear case for including satisfaction as an outcome. Trust and confidence may be appropriate for monitoring, or for reputational incentives. We do not think that there is, at present, a good case for including loyalty as an outcome. We think the case for value for money and fairness is less strong than the case for customer satisfaction.

Table 3: Summary of our conclusions on outcomes

Outcome	Assessment
Satisfaction	🟢
Trust and confidence	🟠
Loyalty	🔴
Value for money	🟠
Fairness	🟠

Source: Economic Insight analysis

4. Measuring outcomes

In this chapter we analyse the most appropriate ways to measure customer satisfaction and trust. This involves both determining which measures are the best proxies for the underlying outcomes of concern, and determining how we make these measures statistically robust in practice.

Our findings:

(i) Surveyed customer satisfaction remains the most appropriate satisfaction measure. There is a case for basing the new incentive mechanism entirely on surveyed satisfaction.

(ii) Some consideration should be given to reducing the amount of time between firms' interactions with customers and surveying customers.

(iii) This could be achieved by changing the channel through which surveys take place.

4.1. Introduction

Our analysis of outcomes suggests that customer satisfaction continues to be the most appropriate outcome to target with financial incentives, while it may also be appropriate to monitor trust and confidence in the water sector. We therefore turn to identifying the most appropriate way of measuring these outcomes. Our analysis suggests that there are two steps in determining the best ways to measure the outcomes of interest.

First we need to determine the best type of measure for the outcome. This refers both to the choice of qualitative measure, for instance, net promoter score versus customer effort score, and to the balance between qualitative and quantitative measures. In practice, this involves determining which measures are most directly correlated with the underlying outcome, and will be credible, trusted and understood by stakeholders.

Secondly, for each type of measure, we need to determine how to make sure that it is, in practice, statistically robust. In practice, this involves considering the following key questions:

» Have the right customers been asked?
» Have the right number of people been asked?
» Were they asked at the right time, using the right channel?
» What are the considerations that determine appropriate governance and funding arrangements?

Overall, therefore, we have three criteria:

» Quality of how the measure reflects the underlying outcome.
» Credibility, trust and understanding of the measure among stakeholders.
» Statistical robustness of the measure.

We first consider how these points apply to customer satisfaction, before then moving onto trust and confidence.

4.2. Measures of satisfaction

This section sets out our assessment of the best ways to measure satisfaction. We first cover qualitative measures, before moving on to discuss the role of quantitative measures.

4.2.1. Qualitative measures

4.2.1.1. Current approach and options for the future

As set out in chapter 2, the current approach to customer satisfaction measures a weighted combination of surveyed customer satisfaction (the qualitative SIM, 75%) and total number of unwanted contacts and complaints (the quantitative SIM, 25%). The qualitative SIM is based on the results of a customer satisfaction survey, in which customers are asked to rate their satisfaction with the service that they received on a 1-5 scale.

When Ofwat consulted on the SIM at PR14, there was significant disagreement as to whether this approach was appropriate. Some stakeholders suggested that the SIM should be based on measures such as net promoter score, customer effort or value for money. Others suggested measures such as the clarity and accuracy of bills and ease of access to information.

Our analysis of measures used in competitive markets, by third parties and by other regulators suggested the following alternative approaches, in addition to retaining surveyed satisfaction levels.

» Net promoter score
» Customer effort score
» Sentiment analysis

Net promoter score and customer effort score could be thought of as being primarily concerned with loyalty and effort respectively. They may, however, be appropriate measures of underlying satisfaction, to the extent that they are correlated with it.

4.2.1.2. Customer satisfaction

Surveyed customer satisfaction remains in wide use among firms in competitive markets. As with the current SIM, customers are generally asked questions along the lines of "Overall, how satisfied or dissatisfied are you with the service you received?". This is generally on the basis of five-point (in the case of the SIM) or six-point scales ranging from very dissatisfied (1) to very (5) or completely satisfied (6).

Surveyed satisfaction is the most direct measure of our variable of interest, and we think that simply asking customers to rate their overall service provides all the information that is needed for regulatory purposes. Indeed, there are advantages to asking *only* about satisfaction rather than particular aspects of the service that customers received, as this reduces the risk of implicitly pre-judging the aspects of service that are important to consumers.

The potential for further innovation in how firms provide customer service means that any measure will need to be sufficiently broad as to reflect new and unforeseen developments. Because it is the most direct measure of satisfaction, and because satisfaction is expressed at a sufficiently high level, surveyed satisfaction meets this requirement.

In view of concerns about convergence, there may be some advantage to extending the scale on which satisfaction is measured. The choice of scale involves trade-offs. Too narrow a scale will make it difficult to distinguish genuine differences in performance across firms. On the other hand, a scale with 'too many' points may introduce 'artificial' variation, and make it difficult for consumers to understand what each point in the scale means.

Firms in competitive markets have sometimes found that "opening up" the top end of the distribution of a five-point scale to a six-point scale can be useful in identifying the very best levels of customer service.

Surveyed satisfaction is also likely to be a credible and trusted measure. It is clearly understood by stakeholders, who have experience of it through the current SIM, and would be understood by consumers and consumer groups. Ensuring that the survey itself is statistically robust will be important in achieving such credibility in practice.

4.2.1.3. Net promoter score

Net promoter score was proposed by Frederick F. Reichheld as a measure that companies could target to drive growth.[6] It is widely used by firms in competitive markets, though there is some evidence that it is falling out of favour. It measures consumers' self-assessed likelihood of recommending a company or service to others. The score itself is based on consumers' answers as to how likely they would recommend a company or product to friends or colleagues, on a scale of 1 to 10. Net promoter score is calculated as the proportion of consumers that answer 9 or 10 (promoters) less the proportion than answers between 0 and 6 (detractors).

The score itself has been controversial. Some researchers found low correlations between net promoter score and revenue, and that it did not outperform consumer satisfaction as a predictor of growth.[7] It is also subject to material fluctuations over time, sometimes without obvious explanation and even while other measures, such as surveyed satisfaction, remain stable.

The likelihood of recommendation is correlated with satisfaction, as shown below, though this is imperfect relative to asking directly about satisfaction. In general, the likelihood of actually recommending a company varies from sector to sector and in industries where recommendations are unlikely in reality, even high performing companies struggle to be rated the 9 or 10 needed to gain high net promoter scores.

---

[6]   *'The One Number You Need to Grow.'* F.F. Reichheld, *Harvard Business Review (2003).*

[7]   *'A Longitudinal Examination of Net Promoter and Firm Revenue Growth.'* T.L. Keiningham, B. Cooil, TW Andreassen, & L. Aksoy, *Journal of Marketing 71 (2007).*

Figure 18: Correlation between UKCSI customer satisfaction measures and net promoter score, water companies and UK sector averages
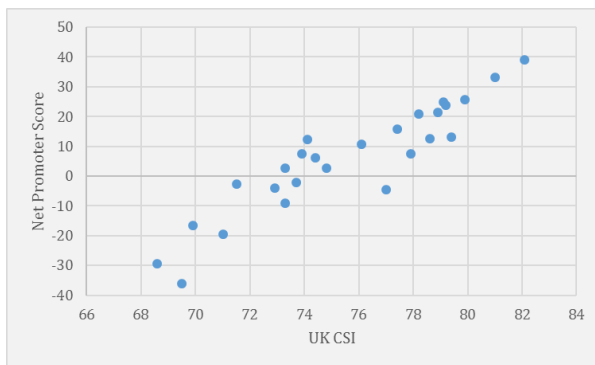
Source: Economic Insight analysis of UKCSI data

This may be particularly problematic in a monopoly situation, in which customers would not expect to make recommendations. Reichheld himself noted that the concept of the likelihood of recommending a product may not be suitable in circumstances in which consumers do not have a choice, due to being too abstract. In such circumstances he suggested alternative formulations, such as 'deserves your loyalty' or 'sets the standard for excellence'.[8]

As it is widely used in competitive markets, net promoter score is in general likely to be credible. There are some issues associated with unexplained variation, however, which mean that it may not perform as highly with respect to credibility as surveyed satisfaction.

Practical example: John Lewis

John Lewis are consistently ranked as having one of the UK's best customer service records, and was ranked in the top ten of all companies in the July 2016 UKCSI report. Again, we note the use of real-time data for measuring a range of customer satisfaction metrics.

Focus

Since 2008, John Lewis have run a bespoke multi-method programme that delivers real-time data on customer feedback accessible through a portal. Surveys conducted to measure service across all business channels.[9]

» Net promoter score (on a 0-10 scale) alongside reasons for that rating (as free text)

» Ratings for the following three aspects of customer service on a 1-5 scale:

  – service at the tills
  – staff made you feel like a valued customer
  – staff available when needed

Data collection methods are shaped to each context in which customers engage with the brand, for example, following online orders, a sample of customers are emailed a link to the survey.

John Lewis also place an emphasis on measuring employee satisfaction as an indicator of likely customer satisfaction.

Rationale

The underlying reason for improving satisfaction is to improve company performance. By measuring customers' likelihood to recommend the brand, they are able to monitor service levels, and if any dips are observed, determine the reasons for this through the open comments given.

How information is used

The real-time measure of satisfaction is used throughout the company, with the NPS score reported at company, area and store level as appropriate. It is used to encourage staff to deliver excellent service and as a diagnostic tool. Quarterly board reports set out how the key drivers of customer experience match with business objectives.

Source: Caroline Thompson Associates research

4.2.1.4. Customer effort score

The idea of the customer effort score is to measure how easy it is for customers to interact with an organisation. It is based on the idea that in many situations consumers do not need to be 'delighted' with top level service. Rather, what customers want is for their interactions with companies to be easy and straightforward.

---

[8] *'On the one number you need to grow, one size doesn't fit all'. D Randall Brandt, Marketing Management (2007).*

[9] www.johnlewisfeedback.com

[10] Many firms in competitive markets monitor their effort score, and it also forms part of third-party measures such as UKCSI.

The effort score is generally measured through a survey that asks for agreement or disagreement with the statement that "the organisation made it easy for me to handle my issue." Answers are normally given on a seven-point scale, from "strongly agree" to "strongly disagree". In some cases, the wording of the question is changed to "How easy was it for you to …". Some companies rephrase the question, for example BT uses "How easy was it to get the help you wanted today?" and use a nine-point scale rather than the original five-point scale. BT then extracts a 'net easy' score in a similar way to net promoter score.

Effort imposes a cost on consumers, and indeed may be more important to them than first-rate service, depending on the context, so in principle appears likely to have a material level of correlation with satisfaction. We note, however, that the data we examined from UKCSI did not show a strong correlation for water companies, though we acknowledge that there are caveats relating to sample sizes.

Effort is also expressed at a lower level than satisfaction or trust, at the level of things that consumers experience, which means that it accords less well with value than directly asking about satisfaction. For instance, suppose that residential retail consumers put primary value on reduced effort, whereas for business customers getting the service right may be most important. In this case effort, would accord better with satisfaction for some customers than for others.

Figure 19: Example outcomes hierarchy for effort
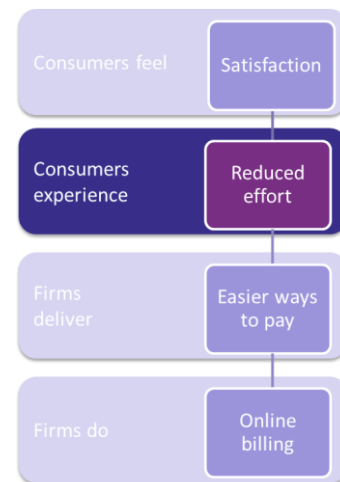
Source: Economic Insight analysis

While it is widely used in competitive markets, and would in general be a credible measure, overall we think that customer effort is less well correlated with our underlying outcome of interest than customer satisfaction. It appears likely that a number of other factors also contribute importantly to customer satisfaction.

4.2.1.5. Sentiment analysis

Rather than using surveys and focus groups, sentiment analysis uses social media to determine how customers feel (possibly in real time). It takes comments from social media about companies and analyses them using computer algorithms to code responses by subject, generally categorising them into positive, negative and neutral subdivisions. Comments are based on customers' own agenda and in their own words, unlike in prompted survey questions.

Sample sizes are often high and results are generally processed quickly. In general, the costs of performing sentiment analysis are concentrated in the initial development of the algorithm, with on-going costs being much more modest.

Sentiment analysis works best for industries that are actively discussed on social media. Analysis would need to be undertaken to see if water companies are sufficiently discussed to make sentiment analysis useful. It is also

---

[10] *'Stop Trying to Delight Your Customers.'* M. Dixon, Harvard Business Review (2012).

unclear how successful sentiment analysis would be in quantifying a particular individual issue such as satisfaction (as opposed to general positive or negative sentiment).

There may be issues around the representativeness of those using social media. Many customers of water companies will not use social media. Ofcom estimates that in 2015 73% of adults with online access used social networking sites.[11] If this subgroup differs in any way from those not using social media then results will be inaccurate for the customer base as a whole. On the other hand, this does mean, that sentiment analysis could be useful for understanding particular customer segments.

The Natural Language Processing algorithms that sentiment analysis uses are also imperfect. Experts estimate that they will only correctly code in 60-80% of cases, and struggle with some of the complexities of language such as sarcasm.

As the new incentive mechanism would apply from 2020, we should bear in mind the possibility of future developments in the extent of social media use for company contact. It may be that the extent of use increases, making it more feasible to generate a representative sample. For example, there is evidence that between 2014 and 2015, the proportion of social media users using their platforms to make a complaint increased from three percent to a quarter. Customers aged 45-54 were most likely to do so, followed closely by 18-24 year olds. [12]

Overall, while sentiment analysis has the potential to be a useful measure, more work would be necessary to establish the representativeness of social media users of the overall population of water consumers, and the accuracy of the algorithms. Further work would also need to be done to establish whether water companies are sufficiently discussed on social media.

4.2.1.6. Other innovative approaches

We examined some of the other innovative approaches that firms in competitive markets use.

» Rant & Rave
» Forrester CX Index
» YouGov brand index
» Brain Juicer tracking
» Research communities and panels.

We provide more detail on these measures in the section 5 of the annex.

Some of the approaches, such as Rant & Rave, are platforms, through which some of the standard customer experience metrics are measured, rather than themselves being measures. Of those that are measures, our assessment is that they are more likely to be suitable for water companies seeking to improve their performance., and are less relevant for promoting innovation through the new incentive mechanism, for one or more of the following reasons.

» They often focus on diagnostic information, rather than determining levels of satisfaction, that can be compared across companies.
» Some are focused on brand perception, which is not an appropriate outcome for the new incentive mechanism.
» The current monopoly situation means that some questions from these measures are inappropriate in the context of water.
» The monopoly situation can also make it complex to compare water company performance with brands in competitive markets.

4.2.1.7. Conclusion

Overall our analysis suggests that surveyed customer satisfaction is the best way of measuring satisfaction. While net promoter score and customer effort score have some merit, there are significant advantages to measuring satisfaction directly. Sentiment analysis has some potential to be a useful measure, but more work is needed to establish

---

[11] Ofcom Media Literacy Tracker

[12] *'Service Goes Social: how organisations can deliver customer service through social media.'* Institute of Customer Service (2015).

whether it is sufficiently accurate, whether it can generate a representative sample and whether water companies are sufficiently discussed on social media.

Table 4: Summary of our conclusions on measuring customer satisfaction

Outcome	Assessment
Customer satisfaction	🟢
Net promoter score	🟠
Customer effort score	🟠
Sentiment analysis	🟠

Source: Economic Insight analysis

4.2.2. Quantitative measures

4.2.2.1. Current approach and options for the future

The quantitative SIM is currently based on the number of complaints and unwanted contacts made to companies, weighted by their seriousness, and then scaled by companies' numbers of connected properties. This score accounts for 25% of the total SIM score.

When Ofwat consulted on the SIM at PR14, companies raised a number of concerns about the quantitative SIM, related to discrepancies in the ways that firms recorded contacts and difficulties in identifying the origin of unwanted contacts. Concern was also expressed that new forms of communication, such as social media and web chat, were not included in the score, and it was suggested that this be corrected.

Other stakeholders, however, have expressed the view that these new media make communications easier for customers. If so, this could both increase the number of contacts and increase customer satisfaction.[13] This would mean that additional contacts were not

necessarily associated with worse customer service performance.

There was also significant disagreement as to the balance between the qualitative and quantitative measures when Ofwat consulted at PR14. Some respondents argued that the quantitative score showed significant variation, meaning that it had more scope for firms to improve on. Others argued that consumers' views should fully determine companies' scores.

Our analysis of the quantitative SIM and stakeholders' concerns suggests the following options:

» Retaining the current quantitative SIM element in the new incentive mechanism (either with a 25% weight as now or a reduced weight).

» Removing the quantitative element entirely.

4.2.2.2. Discussion of options

There are advantages in having a measure of things that consumers do, in addition to measuring their stated preferences and assessments. For instance, knowing that a consumer has made the effort to lodge a complaint gives a strong indication of something going wrong. We also note that firms in competitive markets, across all sectors, and other regulators make use of complaints as a metric.

On the other hand, there are advantages to focusing the new incentive mechanism entirely on measured satisfaction. There is also evidence that very high quality complaint handling can actually lead to increased satisfaction.[14]

Further, the other common performance commitments could be viewed as performing a similar role to the quantitative SIM, by providing an indication of the scale of problems. Improving performance on these performance commitments should lead to fewer complaints, though it should be acknowledged that complaints cover a wider set of issues. Focusing wholly on satisfaction also avoids the issue of

---

[13] *'The future of the Service Incentive Mechanism: A report for Affinity Water.'* Frontier Economics (2016).

[14] *UK Customer Satisfaction Index – The state of customer satisfaction in the UK (The Institute of Customer Service, July 2016) p14*

the regulator putting implicit weight on certain aspects of customer service performance.

We think that there is a valid concern that the quantitative SIM results in perverse incentives. Although making communications easier is likely to be of value to consumers, the quantitative SIM gives firms incentives not to do this because it makes complaints and unwanted contacts more likely.

Given developments in technology, we think that this is likely to decrease in value in the future. As firms could effectively be punished for giving consumers more convenient communication options, we therefore think that the advantages of removing the quantitative element exceed those of retaining a complaints measure.

4.2.2.3. Conclusion

Our view is that the most appropriate option is to remove the quantitative element from the incentive mechanism entirely, so that the focus is entirely on measured satisfaction.

4.3. When and how to survey customer satisfaction

4.3.1. Current approach and options for the future

There are two issues associated with **when** to survey consumers.

» How frequently should surveys take place? While underlying customer attitudes generally do not change rapidly, external events can have a short-term impact on consumers' views. If fieldwork is undertaken during one of these periods it can have a significant effect on the measured outcome, which is nevertheless not sustained. (Though this is less of a concern when measurement takes place across a number of years over a price control.)
» How long after customers' interaction with their companies do surveys take place? Best practice in customer research is to minimise the gap between customers' experiences and their being surveyed.

Options include:

» Increasing the frequency with which qualitative surveys take place.
» Decreasing the amount of time between customers interacting with companies and taking part in the survey.
» Enabling the above by changing the channel through which surveys are conducted.

4.3.2. Discussion of options

The fact that external events can have short-term impacts makes it beneficial to spread fieldwork out at different points of the year to reduce this risk. The current quarterly structure of interviews does provide some spread across the year, but this would be enhanced by using technology to further increase such frequencies.

Our advice is that the longer the delay in consumers giving their response as to their satisfaction, the more likely it is that customers give 'average' responses that do not reflect their actual satisfaction.

To bring the survey for the new incentive mechanism in line with best practice, consideration should be given to shortening the time period between firms' interaction with the customer and customers' surveyed response. This would also contribute to the credibility of the new incentive mechanism.

While this could be done simply by shortening the length of time between contact and telephone surveys taking place, this presents an opportunity for firms to innovate by changing the channel through which data are collected.

For instance, at the end of their interaction with companies, consumers could give immediate responses through text responses to a simple consumer satisfaction question. Given the scope for innovation, there may be advantages to giving companies discretion as to the channel through which responses to a standardised question are gathered. This could allow the development of new ways for consumers to give feedback.

4.3.3. Conclusions

Consideration should be given to reducing the amount of time between interacting with companies and stating their satisfaction. This could be done by changing the channel through

which customers give feedback, for example to allow consumers to give instant feedback via text.

4.4. Including the right customers when measuring satisfaction

4.4.1. Current approach and options for the future

The present qualitative SIM survey is based on a sample that includes only customer contacts. It therefore excludes customers that have not had a direct interaction with their water company. In responses to Ofwat's consultation at PR14, a number of concerns were raised about the narrowness of this sample. For instance, Which? said that customers that have made contact on average represent only 12% of the customer base across companies.

An important development in customer service more generally is the trend towards 'self-service'. Across a range of sectors, consumers are increasingly managing their own accounts, and using company-designed tools to diagnose and solve their own problems. This points to a trend in which increasing numbers of consumers will experience 'indirect' contact with companies, for example through their websites.

Options for the scope of the customer sample include:

» Retaining the current focus on direct customer contacts.

» Increase the scope to include indirect, as well as direct, customer contacts, for example website users.

» Include all customers.

4.4.2. Discussion of options

There are advantages and disadvantages to a wider sample. Including all customers would mean that the sample reflected the entire population, and ensure that incentives reflected their views. At present, incentives do not reflect the views of, for example, users of companies' websites that do not have direct interaction with companies.

Further, if some customers do not contact their companies because they view doing so as pointless, this would not be picked up in the

survey even though it is highly relevant for customer service. Customers may also not have time to contact their companies, or might only provide feedback if they were explicitly asked for it.

On the other hand, customers that have had contact are likely to be best placed to give an assessment of firms' performance. They are more likely to have the information and experience needed to give an accurate assessment of firms' customer service performance.

Measures among a general sample also tend to be stable over time, with changes in service levels taking longer to be observed in overall satisfaction. We expect that if financial incentives are involved, this would reduce the incentive power of the measure.

4.4.3. Conclusions

Overall, widening the sample to include both direct and indirect customer contacts strikes the right balance between achieving incentive power and accommodating relevant views and reflecting changes in consumer behaviour. This would involve adding to direct contacts customers that had interacted with companies using self-serve methods, through social media or by using companies' websites.

Changing the channel through which surveys take place would be useful in getting the views of indirect contacts, such as web users.

4.5. Including the right number of customers when measuring satisfaction

4.5.1. Current approach and options for the future

Quantitative SIM scores are known numbers, and so are not subject to any sampling error. This discussion therefore only applies to the qualitative survey. We note that using new channels to gather SIM data, for example online and through SMS, could make is easier to collect larger samples.

For the purposes of reputational incentives, SIM scores are published on an annual basis. These annual scores are based on a sample size of 800

for the qualitative SIM, collected across the year in four batches of 200.

For the purposes of financial incentives, SIM scores are calculated as an average over the years of the price control. At the end of PR09 this meant that the qualitative component (at the time 50% of SIM scores) was calculated using three years of data, giving an overall sample of 2,400. At the end of PR14, if scores are calculated using five years of data, this will lead to an overall sample of 4,000 for the qualitative component (75% of overall scores).

A number of stakeholders at the PR14 consultation suggested that the sample size of the qualitative SIM needs to be increased both in terms of population and the frequency with which it is collected. Some companies suggested that the sample size should be made bigger by increasing the frequency of the qualitative survey.

Another company noted that small changes in scores have a significant impact on companies' relative positions and argued that statistical noise could be reduced by increasing the sample size to 400 surveys per quarter. Frontier Economics has suggested that sample size be increased to ensure that the accuracy of the results reflect the scale of risks and rewards associated with service performance.[15]

A number of stakeholders raised sample size in the context of convergence. Thames Water and questioned the implications of convergence for penalties and rewards, suggesting that it may be inappropriate to base significant differences in penalties and rewards on what are in fact small differences in scores.

At a high level, options for the number of customers included comprise:

» Maintain the current sample size.
» Increase the current sample size until confidence intervals are narrower.
» For reputational incentives, place heavier caveats to reflect wider confidence intervals from the smaller sample.

4.5.2. Discussion of options

In general, larger samples are better, but gains in accuracy from including more responses tend to decrease as the sample size gets larger. We discussed above that it would be appropriate to maintain a focus on surveyed satisfaction, but increase the speed with which consumers are sampled, which could be achieved by encouraging companies to make innovative use of technology. The latter approach could make it easier to increase sample sizes.

To help us understand the sample sizes that may be appropriate for the future incentive mechanism, we have examined data on the confidence intervals around historical data from the qualitative survey. As we cannot know the spread of future customer responses, this analysis will not be conclusive. It can, however, give a useful indicator as to the sorts of sample size that are likely to be needed in the future.

For financial rewards and penalties five years of data over a price control with an annual sample of 800 would yield an overall sample of 4,000. This is likely to be sufficient to produce reliable estimates and rankings. We note that confidence intervals were fairly narrow around satisfaction scores from the survey data used at PR14, which included fewer observations for each firm (2,400, i.e. 800 over three years).

Figure 20: Confidence intervals around qualitative survey data used at PR14
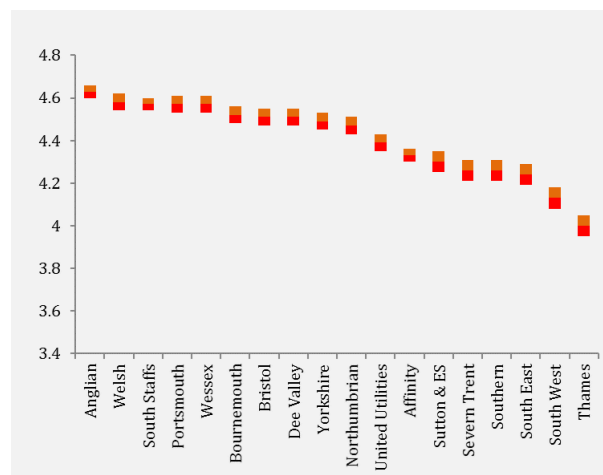
Source: Economic Insight analysis of qualitative SIM data

---

[15] *'The future of the Service Incentive Mechanism: A report for Affinity Water.'* Frontier Economics (2016).

Including more years' data has the potential either to increase or decrease confidence intervals. While the inclusion of more data points would suggest that confidence intervals should fall, the addition of more data could increase the spread of results.

For reputational incentives, sample sizes of 800 are used, which could result in material overlap. For example, on 2015/16 data, if first-ranked Portsmouth Water's 'true' score was at the bottom of its confidence interval, its score could be the same as if ninth-ranked South Staffordshire Water were at the top of its confidence interval.

Figure 21: 95% confidence intervals around 2015/16 qualitative survey scores
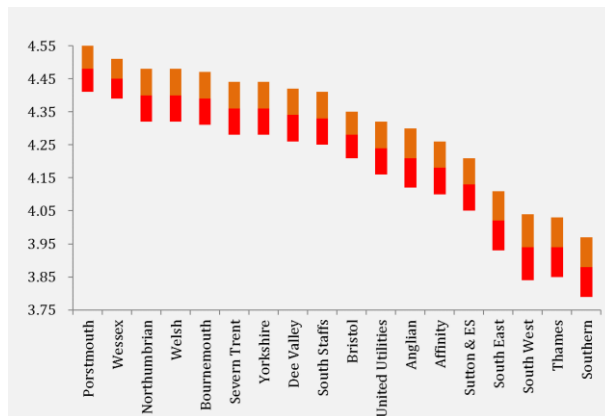
Source: Economic Insight analysis of qualitative SIM data

This may not arise as an issue if innovative use of technology allows for sample sizes to be easily increased in any case. Assuming standard deviations are constant, increasing the sample size to 2,000 could reduce confidence intervals around survey scores from 0.06-0.10 to 0.04-0.06. This is shown in the figure below, which shows how confidence intervals would vary with sample size for a number of companies (including those with the widest and narrowest sample size).

Figure 22: Effect of increasing sample size on confidence intervals around qualitative SIM surveys
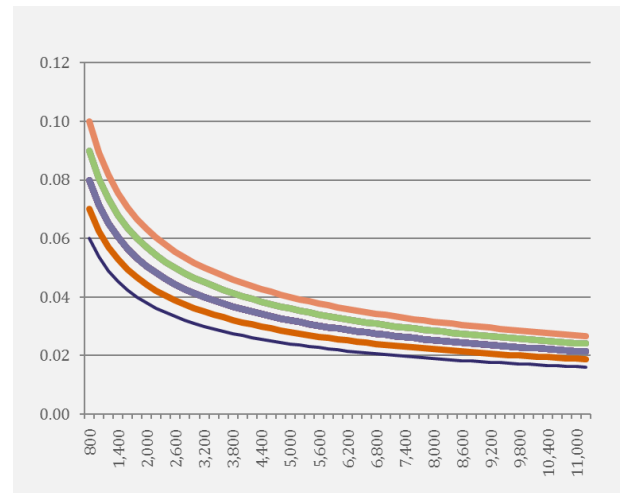
Source: Economic Insight analysis of qualitative SIM data

If sample sizes were to remain at 800, the extent of overlap in scores suggests that there is a case for more heavily caveating annual scores, for example by using bands or explicitly providing confidence intervals.

4.5.3. Conclusions

The qualitative survey sample is sufficiently large that differences in the overall SIM score, and therefore in rewards and penalties will reflect differences in performance. Innovative use of technology may enable further increases. This could also address concerns about confidence around firms' annual scores, though providing sufficient caveats could also provide a solution.

4.6. Measuring trust and confidence

As set out in the preceding chapter, we think that trust and confidence may be appropriate for **monitoring**, and possibly reputational incentives, though they are unlikely to be appropriate for direct financial incentives. We now consider how the measurement issues that we examined above for customer satisfaction apply to trust and confidence. Apart from matters around the choice of measures and the choice of customers, the other issues surrounding trust measurement are the same as for customer satisfaction, and we do not discuss them in more detail.

4.6.1. Which measures?

The choice of trust measure depends to a large extent on the extent to which Ofwat would wish to monitor trust in the sector as a whole or in individual companies. Some off-the-shelf trust measures are available, including as part of UKCSI and from Which? The UKCSI survey only surveys customers of some water companies, so currently can only give an incomplete view of trust in either companies or the sector. The Which? survey asks about trust in the sector as a whole, which could be of some use to Ofwat.

We note, however, that there are discrepancies between UKCSI and Which? survey data. When customers are asked about trust in the sector as a whole in the Which? data, and specifically for water supplied to their homes, water scores highly. On UKCSI data, when customers are asked about individual companies, water companies in general have low scores relative to other sector averages.

Figure 23: Which? trust data for industries
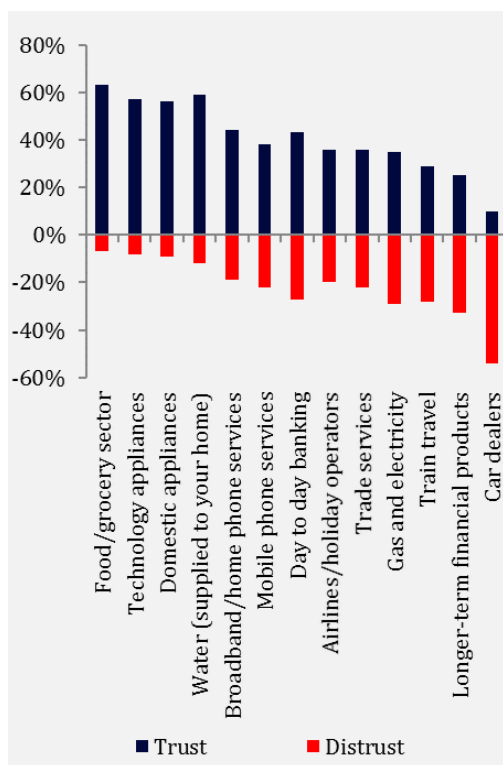
Source: Economic Insight analysis of Which? data

Figure 24: UKCSI Trust scores for Water companies and UK sector averages
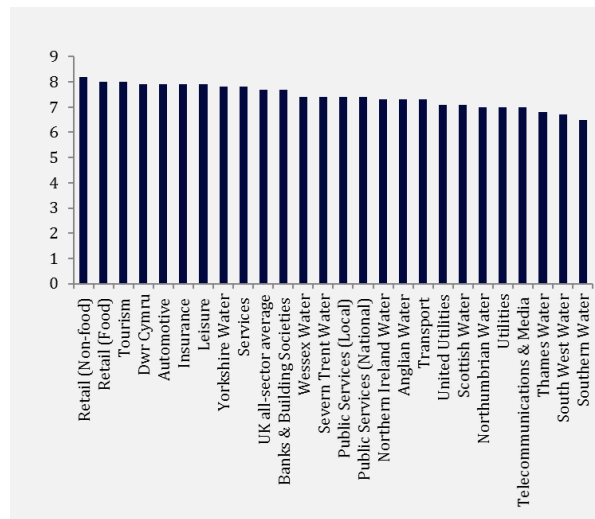
Source: Economic Insight analysis of UKCSI data

As neither off-the-shelf measure gives a separate score for all companies, getting company-level data on trust will require bespoke research. This has been increasingly common in the food retail sector in recent years. Trust is usually measured through statements such as 'a reliable, trustworthy company', 'a company I can trust', 'a company I can trust to do the right thing' or 'a reputable company'

A variety of scales and labels are used, usually at the preference of research agencies. Traditionally, five-point scales were used with each point labelled, but a variety of others are in use, including binary 'trust/distrust' questions and scales of 1-10 (though in this case responses often converge around the 7-9 range).

4.6.2. Which customers?

Unlike customer satisfaction, we could think of trust as being appropriate both for all consumers and for the narrower set of consumers that have experienced direct interaction. As trust reflects customers' expectations of company performance, it is relevant for those that have not had an interaction with their company.

Alternatively, trust could be framed in the context of the quality of companies' service at the point of interaction with consumers, in the same way as customer satisfaction.

4.7. Governance and funding

'Governance' refers to the processes by which the gathering of data for the incentive mechanism, including the conduct of consumer surveys, is managed. Funding refers to how this data collection is paid for.

For the SIM in its current form, firms collect data on the quantitative component of the SIM and calculate scores, which they submit to Ofwat. A third party carries out the qualitative survey, which is procured and funded (through firms' licence fees) by Ofwat. Ofwat then calculates overall scores from the quantitative and qualitative data.

4.7.1. Governance arrangements

At a high level, governance covers both the collection of data and the management of the data collection process. Management of the process could be undertaken by Ofwat or by individual firms, while the collection of data itself could be undertaken directly by firms, or by a third party. In the case of third party data collection, there is also the issue of the choice between bespoke and off-the-shelf research.

4.7.1.1. Managing the process

There are advantages and disadvantages to each of these options. If Ofwat undertook the management of the process for the new incentive mechanism, as they currently do for the qualitative SIM, this would have the advantage that it would be perceived to be independent. There are also likely to be efficiency advantages if data collection was to be undertaken by Ofwat or a single third party, as this would save the costs of multiple firms engaging with the same organisation or multiple organisations.

On the other hand, having Ofwat manage the process could reduce firms' sense of ownership over satisfaction measurement. There could also be some circumstances in which it would be more cost-effective for firms to manage the process directly.

Having firms manage the process would not have the advantage of being seen as independent, though it would give firms greater flexibility. For instance, they could decide themselves whether to engage third parties to collect data or whether to do it themselves. Alternatively, firms could be required to appoint independent third parties. This would achieve independence and allow flexibility in the management of the process.

4.7.1.2. Data collection

Having firms collect data may not be seen as independent, though Ofwat could establish clear rules to encourage confidence in the data collection process. This should be possible, as similar issues apply to firms' collection of data for the quantitative part of the current SIM. Having firms collect data could also give them scope to innovate, for instance asking additional questions that would give them useful diagnostic information.

Engaging a third party would have the advantage of being perceived as independent. Third parties are also likely to have more expertise in data collection, though we note that if firms are responsible for the management of the process they could always choose whether or not to collect data themselves or to engage a third party to do so.

Our recommendation of exploring quicker ways of gathering customers' views has implications for the way that data is collected. As firms are, by definition, already in contact with customers, engaging a third party to undertake data collection could be unnecessarily costly, when firms could build the collection of data into their systems. There may therefore be a trade-off between having more immediate customer service measurement and the centralisation of the governance process.

4.7.1.3. Bespoke versus off-the-shelf research

If a third party was to be used for data collection, the question arises as to whether it is better to use **bespoke** or **off-the-shelf** research. **Bespoke research** is conducted for a specific client (or clients) with the questionnaire, sample, research approach, and timings all designed to best meet the specific research objectives. It is by far the most commonly used approach, as in business most research objectives are specific to that company, at that time.

'Off-the-Shelf' research solutions fall into two types:

» A multi-buyer survey run by an independent organisation who then sells it to multiple client companies. An example of this would be the Institute of Customer Service's UKCSI.

» A research 'system' which is typically owned by one of the large research agencies and employs a standard methodology. Clients then individually buy into this system, but can then compare their results to the normative data held by the research agency. An example is the Forrester CX Index (see the section 5 of the annex).

Bespoke research has the advantage that the sample, methodology, questionnaire, organisations covered and timing are all fully tailored to research objectives. On the other hand, they tend to be more expensive than multi-buyer surveys. Using bespoke measures may make it more difficult to benchmark, even if questions are 'standard' such as trust, due to methodological differences.

Multi-buyer surveys are the cheapest bespoke option. They deliver the ability to benchmark against other organisations included in the survey and are generally trusted and perceived to be independent. On the other hand, they give no direct control over sample, methodology, questionnaire, organisations included or survey timing.

Choosing between bespoke and off-the-shelf options requires trading off the cost advantages of off-the-shelf measures with the extent of control over methodology and comprehensive coverage of companies that a bespoke measure would bring.

In theory, satisfaction could be measured through bespoke or off-the-shelf measures. In practice, however, there are currently no off-the-shelf research systems that include all water companies. Multi-buyer options may find it difficult to find a large enough sample of consumers that have had sufficiently recent contact, so in practice it is likely that one would have to use a bespoke measure to monitor satisfaction. (One could pay to increase the sample size, though this would reduce the cost-effectiveness of multi-buyer surveys.)

As we do not argue that explicit financial incentives should be attached to trust and confidence, there are fewer advantages to having explicit control over the associated methodology. This suggests that an off-the-shelf measure could be appropriate.

4.7.2. Funding arrangements

Options for funding are more limited than governance options, ranging from having firms incur the costs of data collection directly to having Ofwat fund them centrally through firms' licence fees.

In some circumstances it may be advantageous for firms to incur costs directly. This would be the case if there was scope to innovate and reduce costs. Having firms incur the costs of data collection directly give them incentives to minimise costs.

In other circumstances, it may be better for Ofwat to incur costs directly. This has the advantage of simplicity and could avoid unnecessary duplication. For instance, Ofwat would have to procure the survey once, as opposed to 18 water companies. This would include situations in which a single third party was to carry out data collection for a bespoke measure across all firms, or if an off-the-shelf trust measure was to be used.

4.7.3. Conclusion

Governance and funding options vary according to the extent to which they are centralised. Having Ofwat take responsibility for managing the data collection process has the advantage that it would be perceived as independent and could have efficiency advantages. On the other hand, having companies take responsibility could increase flexibility and encourage them to take ownership of the satisfaction measurement process.

Likewise, having firms incur costs directly has some incentive advantages, though in other cases it may be appropriate for Ofwat to fund costs through licence fees in order to avoid unnecessary duplication.

Table 5: Governance and funding considerations

		Advantages	Disadvantages
Management	Ofwat	Perceived as independent; reduces duplication	Reduces firms' sense of ownership; could introduce a 'middleman'
	Firms	Firms have greater flexibility, possibly more efficient	Not seen as independent
Data collection	Third party	Perceived as independent; greater expertise	Less firm ownership
	Firms	Firms are already in contact with customers; give firms scope to innovate with respect to diagnostics	Not seen as independent
Survey type	Bespoke	Specifics of survey fully tailored to research needs	Expensive
	Multi-buyer	Inexpensive, trusted	No direct control over survey design; difficulties in achieving a large enough sample
Funding	Ofwat	Avoids duplication	Does not give firm-level cost minimisation incentives
	Firms	May have positive incentive properties	Can lead to duplication

Source: Economic Insight analysis

4.8. Conclusions

Our analysis of potential measures suggests the following conclusions.

» We think that **surveyed customer satisfaction** remains the most appropriate satisfaction measure.

» There is a case for making the new incentive mechanism **entirely qualitative.**

» It remains appropriate to focus satisfaction measurement only on customers that have had **direct interaction** with water companies.

» The **sample size** of the existing surveys is appropriate for setting financial incentives, though there may be a need to **caveat annual scores** more clearly.

» There are issues around the **length of time** between consumers interacting with companies and the subsequent assessment of satisfaction. Reducing this time could be done by changing the survey channel, for instance by asking customers for instant feedback.

» There are advantages and disadvantages to the various options for **governance and funding**. At a high level, more centralised options have the advantage of being seen as independent, but reduce firms' flexibility and sense of ownership.

5. Benchmarking outcomes

We considered the options for using evidence on performance in other sectors to determine targets for water companies' customer service levels. We then analysed the advantages and disadvantages of using this evidence, against retaining an approach in which companies were benchmarked only against each other.

Our analysis suggests:

(i) There is a strong case for having wider benchmarks.

(ii) It is best to use benchmarking to inform the targets that firms are set, rather than benchmarking in a mechanistic manner.

At present, no use is made within the design or operation of SIM of customer service performance delivered by companies outside of the water sector. Ofwat therefore asked us to consider:

» the options for using cross-sector evidence in new incentive mechanism; and

» the advantages and disadvantages of doing so.

5.1. Options for using cross-sector evidence in the future incentive mechanism

There are various ways of using cross-sector evidence in the design or operation of incentive mechanism, namely:

» First, to improve our understanding of the outcomes that matter to customers and the options for measuring them.

» Second, to improve our understanding of the costs and benefits of delivering different outcome levels.

» Third, to improve our understanding of what level of customer service performance might be achievable "*today*" with the budget companies have through the price control – and hence set an appropriate incentive structure to encourage *catch-up*.

» Fourth, to improve our understanding of what level of customer service performance improvement might be achievable "*over the duration of the price control*" – and hence set an appropriate incentive structure to encourage *frontier shift*.

We focus on the third and fourth options in this section. In relation to both options, the main reason for using cross-sector evidence is to increase the quality of information available to Ofwat and others necessary for understanding the level and rate of improvement of customer service performance that water companies could and should be challenged to deliver during the price control.

In particular, given that the objective of incentive mechanism is to mimic the incentives that firms would face in a "competitive market", there is a good argument for looking at what firms in comparable competitive markets actually deliver.

Indeed, using cross-sector information could help test a legitimate concern that – because of the inherent limitations of regulation compared to competition – water companies "under-deliver" customer service (in terms of its level and/or rate of improvement) compared to what might be achievable with the budget they have. This would make the case for Ofwat to consider setting incentive rates in a way that would encourage bigger and/or faster improvements in customer service performance than at present.

One way of achieving this would be to use the cross-sector information to set a new threshold or benchmark at which penalties and rewards would apply. There are various options as to exactly "how" this information is used and, in particular, the weight that would be attached to it compared to water sector-specific evidence. These options include but are not limited to:

» Ofwat could determine the threshold or benchmark <u>only with reference</u> to the customer service performance it observes in other sectors. For example, it could decide that water companies would only receive rewards if they achieve the customer service performance of the "top X" or "upper quartile" firms as measured by a third-party provider, such as UKCSI.

» Alternatively, Ofwat could determine the threshold or benchmark <u>with reference to a range of evidence</u>, including but not exclusively the customer performance it observes in other sectors. For example, it could set the threshold based on the average of the upper quartile performance of the water companies and the upper quartile performance of companies in other sectors. It could also use cross-sector evidence to inform whether (say) it would be appropriate to set the threshold or benchmark at the upper quartile of water company performance or a higher (e.g. upper quintile), but without a direct "arithmetic" link to cross-sector performance.

Below we consider the advantages and disadvantages of using cross-sector evidence at all, and using it exclusively or using it alongside other water sector-specific evidence.

5.2. Advantages and disadvantages of using cross-sector evidence in the new incentive mechanism

5.2.1. Using – or not using – cross-sector evidence

The main advantage of using cross-sector evidence is noted above i.e. it potentially provides extra information on the appropriate threshold or benchmark at which penalties and rewards should become payable.

The main challenge associated with using cross-sector information relates to comparability, namely the marginal willingness to pay and costs of delivering customer service may vary between firms and sectors.

Whether the advantages would offset the disadvantages is likely to turn on four factors.

» The extent to which the mechanism provides incentives for firms to reach and extend frontier service.
» The extent to which other firms and sectors are comparable to the water sector.
» The risk of unintended consequences.
» The availability of comparable metrics in other sectors.

5.2.1.1. Incentives

Even without using cross-sector information, the SIM may provide incentives for firms to catch-up and shift the frontier of customer service. The current SIM is designed in a way that has the potential to encourage frontier shift if the incentives rates are right – that is, the frontier water company faces a risk that it will no longer be the frontier water company if other companies catch-up and outperform it. The size of this risk depends on how easy it is for other firms to leapfrog the frontier water company.

However, supposing that it were difficult for other companies to contest the frontier company, it may not need to "work hard" to retain its position – and this could reduce the rate of frontier shift in customer service in the water sector. We note that in some ways this concern is the opposite of the concern about convergence. This suggests that if scores are too far apart, then the incentive mechanism may lose some of its incentive powers.

Our analysis of SIM data suggests that firms' SIM scores are close enough that most companies would have a reasonable expectation that it would be possible to overtake firms with similar scores. This can be seen in the figure below which shows firms' annual SIM scores over time.

Figure 25: Annual SIM scores over time (not adjusted for change in methodology)
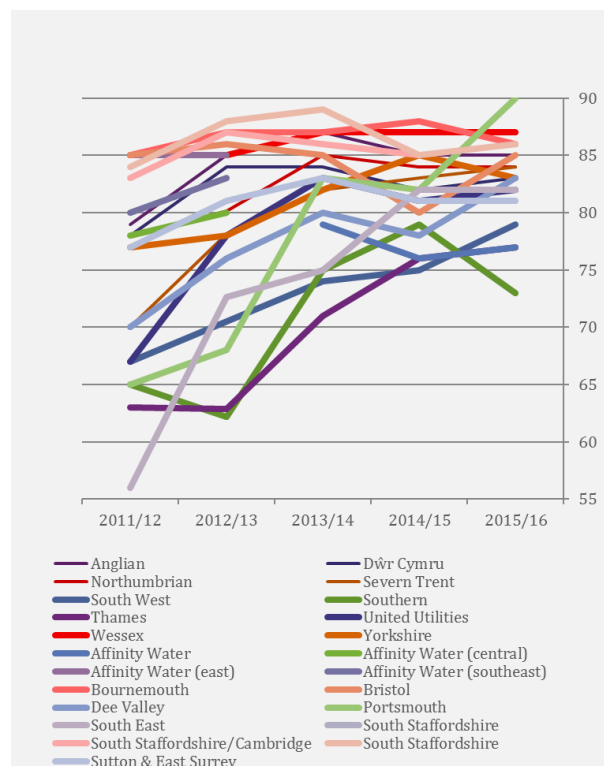
Source: Economic Insight analysis of SIM data

In addition to convergence in general, the highest scores have remained in the region of 85 to 90, although a score of 90 was only achieved this year. We note that this trend may reflect the structure of SIM incentives, which are asymmetric and mean that rewards for high performance are lower than penalties for poor performance.

The closeness of scores at the frontier can be seen in the table below, which shows the scores of the first, second and third highest ranked firms over time. At most three SIM points separate the firms. We can also see that there is considerable year-to-year movement.

Table 6: Highest SIM scores in each year

Year	First		Second		Third	
2015/16	Portsmouth	90	Wessex	87	Bournemouth	86
2014/15	Bournemouth	88	Wessex	87	South Staffs South Staffs (Cambridge) Yorkshire Anglian	85
2013/14	South Staffordshire	89	Bournemouth Wessex	87	-	-
2012/13	South Staffordshire	88	Bournemouth South Staffs (Cambridge)	87	-	-
2011/12	Bournemouth Bristol Wessex Affinity (East)	85	-	-	-	-

Source: Economic Insight Analysis of SIM data

5.2.1.2. Comparability of other sectors

If wider benchmarks are to be used, the extent to which other sectors provide meaningful comparisons for water depends on two factors:

» The extent to which consumers in other sectors place similar valuations on customer service performance as consumers in the water sector.

» The extent to which firms in other sectors incur similar costs in delivering customer service outcomes.

Getting targets right is important. If the other sectors used as benchmarks are not comparable, it may be that the target performance levels that are set for firms are too high or low. In view of the evidence we set out below, which suggests that consumers place less value on high customer service, it may be the balance of risks points to setting too demanding targets. This could result in there being no incentive power, since the only possible outcome would be that the firm would receive a punishment, which would occur with 100% probability.

There are other risks associated with setting benchmarks too high. In order to meet such targets, incentive rates have to be set such that customers are obliged to pay more than they would be willing to in a competitive market, creating deadweight loss.

Alternatively, even if the outcomes were achievable, it could be that allowed revenues were not increased sufficiently to deliver firms funding to deliver them. It could also be that the targets set result in deadweight loss, with costs in excess of benefits. This would occur if the marginal benefits to consumers (i.e. valuations) were lower in water compared to other industries.

For example, evidence from UKCSI suggests that fewer customer place value on high customer service at a higher price in utilities than in other sectors.

Figure 26: Percentage of customers that value high customer service, even at a higher price
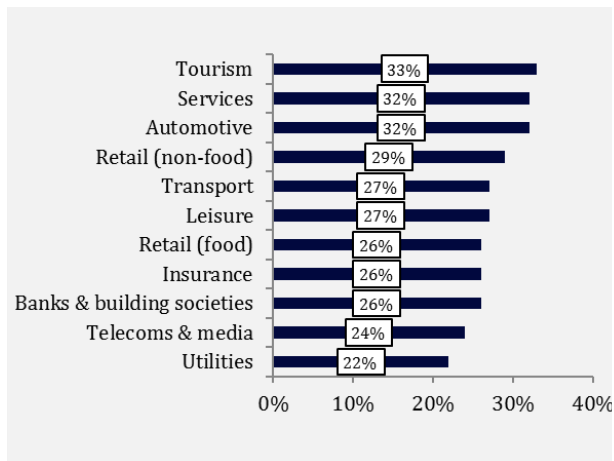
Source: Economic Insight analysis of UKCSI data

5.2.1.3. Comparability of individual firms

Third, and relatedly, the risk of "unintended consequences" that could arise due to variation in performance of individual companies or sectors, depending on how benchmarking worked in practice. There are particular risks with benchmarking with respect to individual firms, as both their relative and absolute performance can vary significantly over time. Benchmarking against wider sectors has the potential to mitigate such risks.

The risks associated with benchmarking against particular firms can be seen when looking at the top five performing firms over the past three years on the UKCSI.

Table 7: Top five performing firms in July UKCSI

	2016			2015			2014		
Rank	Company	Score	2015 score & rank	Company	Score	2014 score & rank	Company	Score	2013 score & rank
1	Amazon	87.2	(86.3, 2)	First Direct	87	(86.3, 3)	John Lewis	87.7	(90.8, 1)
2	Wilko	84.8	(79.8, 45)	Amazon	86.3	(87.6, 2)	Amazon	87.6	(89.0, 2)
3	Waitrose	84.6	(84.5, 6)	John Lewis	86.2	(87.7, 1)	First Direct	86.3	(87.2, 4)
4	M&S Food	84.6	(82.8, 12)	LoveFilm	84.8	(76.6, 111)	Tesco Mobile	85.8	(82.6, 31)
5	Land Rover	84.6	(79.1, 57)	Specsavers	84.7	(84.4, 7)	M&S Food	85.6	(86.7, 5)

Source: Economic Insight analysis of UKCSI data

Amazon, Waitrose and M&S Food consistently perform highly, whereas other firms move around significantly. For instance, Land Rover was in the top five in 2016, but ranked 57th in the precious year, while Love Film moved from 111th to 4th in 2015. Such significant movements in performance could have unintended consequences, for example causing firms to be rewarded as a consequence of an individual benchmark happening to perform poorly, or punished in the event that a benchmark performs unusually and unrepresentatively well.

This risk is partly, but not fully, mitigated when considering benchmarking against other sectors. While this would average out some of the variation in individual companies' scores, it does not rule out the possibility that particular sectors as a whole could undergo changes that were irrelevant for the water company frontier. For instance, an industry-wide scandal could affect a sector and lead to deteriorations in satisfaction.

To some extent this risk can be mitigated by the way in which such cross-sectoral information is used. For example, if the evidence is used alongside water sector-specific evidence, then there is less risk of being "misled" by

information on an individual firm or other sector's performance.

Overall, this suggests that benchmarking against individual companies should be avoided. It also suggests that cross-sector information should be used in a considered manner.

5.2.1.4. Availability of comparable metrics

Fourth, the availability of robust and comparable metrics of customer service performance in other sectors is an important determinant as to whether the advantages of benchmarking exceed its disadvantages. UKCSI provides information on customer satisfaction. Any comparisons made between UKCSI and the new incentive mechanism would need to be sure to take into account any methodological differences between the two.

5.2.2. Using cross-sector evidence exclusively or alongside other (water sector) evidence

The advantages and disadvantages of these broad options are very closely related to the four factors set out above. That is, using a mix of evidence is likely to be a sensible approach when it is uncertain how comparable different sectors are to one another and/or where the risk of unintended consequences is high.

5.3. Conclusions

There are good arguments for looking at what level of customer service performance firms in other sectors achieve when setting the threshold or benchmark at which penalties and rewards are applied.

Because there is and will be uncertainty over how comparable different sectors are to the water sector, cross-sector evidence should be used in a considered manner and alongside other water sector-specific evidence. This uncertainty also points to caution around linking incentives to cross-sector performance in an arithmetic way.

6. Annex

This annex presents additional supporting material.

Section 1 details background to the SIM.

Section 2 examines the complementarity of the current SIM with other common performance commitments.

Section 3 provides additional technical details on customer service incentives under monopoly.

Section 4 provides further information on outcomes frameworks.

Section 5 provides more detail on measuring customer service.

6.1. Background to the SIM

This section presents additional detailed background material on the SIM since it was introduced at PR09.

6.1.1. Original objectives of the SIM

Ofwat introduced the SIM at PR09 to provide incentives to firms to improve service quality and value for money. The SIM replaced the Overall Performance Assessment (OPA), which had been in place since the 1999 price review. The OPA allowed firms to charge customers more or less than they otherwise would have done, depending on the scores that they achieved, which were based on factors such as reliability and response times.

At PR09, Ofwat set out a number of grounds for moving from the OPA to the SIM.[16]

» There had been significant numbers of complaints about companies' quality of service (6% of consumers were not satisfied with basic aspects of service).

» The OPA's focus was on reliability and response times, rather than quality.

» Most companies' performances on OPA criteria had reached acceptable levels, so cross-company comparisons had become less effective at driving improvements.

» By focusing on specifically defined performance standards, the OPA did not award innovation (and may have discouraged it).

» There were large differences in consumer experiences, with satisfaction for individual companies ranging from 60% to 90%.

» The OPA was no longer needed to provide comprehensive performance information.

» The broader performance measures in the SIM gave Ofwat more flexibility to review the need for the various regulatory information requirements that it had in place.

Ofwat further explained that the purpose of the SIM was to incentivise companies to *get things right the first time* and, in the event that things do not go well first time, resolve complaints

quickly and effectively. This was seen as a way of reducing costs and therefore generating savings for companies and consumers through lower bills. Ofwat considered that it was <u>not</u> intended as a safeguard for basic service levels.

Ofwat also said that it saw the SIM as continuing with the *reputational incentive* that the OPA provided, with a comparative assessment of companies' performance to be published. Added to this was a financial incentive, based on the SIM performance measure. Rewards and penalties were to be applied to prices from the start of the following price control period (i.e. PR14).

Ofwat settled on a range of potential financial rewards and penalties, equivalent to those that previously applied to the OPA, with firms being allowed to charge between 1% less, and 0.5% more, than they otherwise would have been able to. Firms' SIM performances between 2011/12 and 2013/14 would be compared, and price adjustments implemented accordingly. (Using data from 2011/12 was intended to allow Ofwat time to ensure that the measure was robust).

Ofwat selected two ways of measuring consumer experience within the SIM:

» The **quantitative** measure covered the number of complaints and telephone contacts that companies received, weighted for their place in the complaints process. Abandoned calls, calls when all lines were busy, and unwanted telephone contacts, were given a weighting of 1; while written complaints were multiplied by 5; escalated written complaints by 100; and complaints investigated by the Consumer Council for Water (CCWater) by 1,000. Email contacts were not included, as they only accounted for around 4% of total contacts and would have introduced additional complexity to the measure. To control for companies' sizes and differences between WaSCs and WoCs, Ofwat normalised the quantitative score using the number of connected properties.

» The **qualitative measure** used a survey of consumers that had had direct contact to

---

16   *'Putting water consumers first – the service incentive mechanism.'* Ofwat (2010).

determine satisfaction. The survey was designed to take individuals through their experience of dealing with the company from first contact to the point at which the issue was resolved. The score was derived from consumers' overall satisfaction, which they were asked to rate on a scale of 1 (very dissatisfied) to 5 (very satisfied). Weighted average satisfaction scores were then calculated, with 50% from billing contacts and 50% from operational contacts.

To determine how companies' scores would affect their financial incentives, Ofwat decided to use performance bands, based on companies' SIM scores, rewarding those *above average* and penalising those *below average*. This reflected the idea that companies would be 'competing' with each other to receive rewards and avoid penalties.

6.1.2. Firms' 2010-2015 SIM performance and related rewards and penalties

Ofwat set financial rewards and penalties for firms' performance in 2010-2015 on the basis of the SIM scores for the years 2011/12 to 2013/14. Three-year average SIM scores over the period are shown in Figure 27, alongside the firms' relative performances in terms of standard deviations from the mean.

Figure 27: Companies' three-year average SIM scores, 2011/12-2013/14
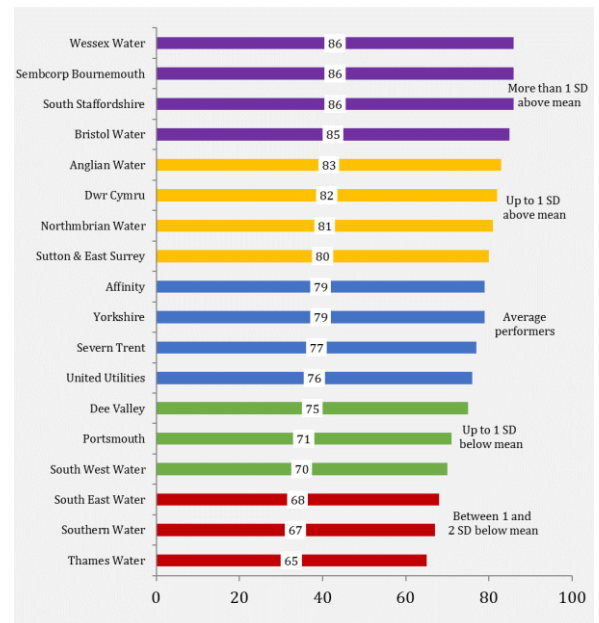
Source: Final price control determination notice: policy chapter A4 – reconciling 2010-15 performance

At PR14, Ofwat did not make its methodology for relating scores to financial rewards and penalties explicit until it had finalised the implications for total allowed revenues, on the grounds that revealing the results to companies could dampen the incentive properties of the SIM. It did, nevertheless, make clear that it would be based on firms' position relative to the industry average and provided firms with a broad indication of the potential value at stake.[17]

Ultimately, at the Final Determination, the relationship between firms' penalties and rewards, and their position relative to 'average' performance, was broadly in line with Ofwat's prior indications, as shown in Figure 28 below.

---

[17] *'Setting prince controls for 2015-20 – final methodology and expectations for companies' business plans.'* Ofwat (2014).

Figure 28: Relationship between financial incentives and SIM scores over 2010-15
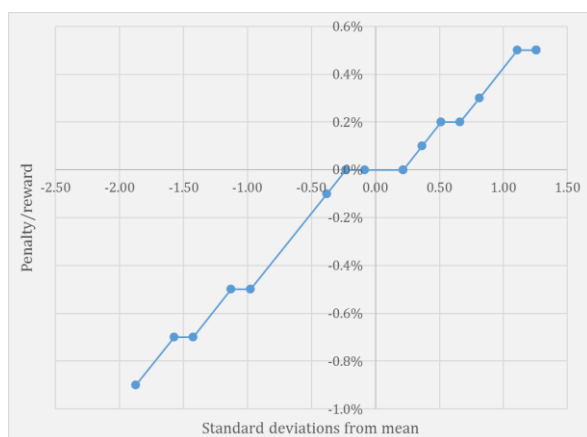
Source: Economic Insight analysis of Final price control determination notice: policy chapter A4 – reconciling 2010-15 performance

Ofwat had already made clear that penalties and rewards would be in the range of -1.0% and +0.5% of what companies would otherwise have been able to charge. As it was also clear that penalties and rewards were to be based on (publically available) comparative performance data, realised rewards and penalties were likely to have broadly matched investors' expectations.

6.1.3. Changes to the SIM at PR14

Ofwat undertook a review of the SIM at PR14, which ultimately led Ofwat to conclude that the SIM should be retained on the grounds that it had been successful in driving improvements in customer service over time. This took place in the broader context of a move from setting companies an overall price control, to setting separate controls for wholesale water and wastewater; and for retail household and non-household. As a result, Ofwat decided that the reputational and financial incentives of the SIM *would apply to household retail controls in England and household and non-household retail controls in Wales.* For the non-household control in England, incentives would be reputational only. The scope of Ofwat's review of the SIM at PR14 included consideration of the following issues:

» Whether the SIM should be used **to provide incentives for the best performing companies, less well performing companies, or a combination of the two.** Ofwat concluded that the focus should remain

on driving improvements among the lowest performing firms, but that the SIM should still encourage the best performers to maintain or improve their position.

» **Whether to maintain an asymmetric structure,** with greater downside risk than upside reward, or move to a symmetric structure. Most respondents favoured the asymmetric structure, which Ofwat maintained. Ofwat also considered whether to change the magnitude of the financial incentive, but decided to maintain the status quo on the grounds that it had been successful in driving improvements. This was, however, expressed *as a percentage of household retail revenues.*

» **Whether to move to a system based on absolute performance levels.** Ofwat decided to maintain the existing system for determining rewards and penalties, using firms' *comparative performance.*

» **The weighting between quantitative and qualitative measures.** Ofwat decided that the SIM should place more weight on the qualitative measures; and so decided to switch the balance between quantitative and qualitative measures from the then equal weighting to a 75:25 split in favour of the qualitative measure.

» **The detail of how quantitative and qualitative measures were calculated.** Ofwat decided to make changes to way it calculated both the quantitative and qualitative measures. For households in England, it removed *abandoned/busy calls* from the measure, due to the associated burden of data collection. Owfat also decided to shorten the qualitative survey and removed the "prior notification" to companies that the survey was due to take place.

Under the prevailing approach, firms' SIM performance over 2015-20 will determine penalties and rewards that will be calculated at PR19 and applied over 2020-25.

6.2. Complementarity of the current SIM

We examined the extent of complementarity between the existing SIM and Ofwat's other common performance commitments. These

measure company performance using operational data, and are therefore conceptually distinct from the service focus of the SIM. There may, however, be scope for overlap, as firms' operational performance is likely to inform how consumers perceive the service levels that firms provide.

6.2.1. Complementarity

There is no obvious way to think about complementarity in the context of the SIM and other performance commitments. On the one hand, complementarity suggests that the SIM and performance commitments would "work together" to provide incentives for the same outcome. On the other hand, we could think of complementarity as meaning that the SIM and the performance commitments incentivise different outcomes, but work together to encourage firms to provide a combination of outcomes that consumers care about.

If particular performance commitments are strongly correlated with the SIM, then it is possible that firms' incentives are duplicated. This can be deliberate, but if not it could lead to firms having incentives to concentrate on particular outcomes at the expense of others, in a way that was not originally intended. This could lead to improvements in performance that exceeded consumers' willingness to pay for them. Alternatively, if the SIM and other performance commitments have a negative correlation, this could lead to intended incentives being reduced.

Given this uncertainty as to what complementarity means in any one concept, we examine the extent of overlap between the SIM and the common performance commitments, in terms of what they measure, and then analyse the extent to which this overlap could be considered complementary or not.

Although the common performance commitments will not necessarily remain the same from PR19, it is likely that there will continue to be incentives based on operational data. This analysis will therefore be useful when considering incentives as a complete package.

6.2.1.1. The SIM and the common performance commitments

Aside from the SIM and leakage (for which data is not normalised) there are five common performance commitments, three of which relate to water and apply to all firms, with the other two relating to sewerage and applying only to WaSCs.

» **Internal sewer flooding**: Number of incidents of internal sewer flooding for properties that have flooded within the last ten years per 1,000 properties.

» **Mean zonal compliance:** Average compliance rate for 39 water quality parameters (including colour, odour, taste and chemical content) from regulatory sampling programme.

» **Category 3 pollution incidents:** Incidents with potential for negative impact on public confidence in the water supply requiring a detailed investigation and assessment of the event by a warranted Inspector.

» **Water supply interruptions:** Number of minutes lost due to water supply interruptions for three hours or longer, per property served.

» **Water quality contacts:** The total number of complaints about discolouration, taste and odour per year.

The SIM comprises four distinct parts – the water, sewerage and billing aspects of the qualitative SIM, and the quantitative SIM. As the quantitative SIM relates to contacts in general, there is scope for overlap with any of the performance commitments, though in practice it may seem doubtful that zonal compliance would drive contacts except in extreme cases. The billing part of the qualitative SIM, on the other hand, has no obvious overlap. This leaves the water and sewerage parts of the qualitative SIM, which have some potential for overlap with the relevant performance commitments.

Table 8: Potential for overlap between the SIM and the common performance commitments

Commitment	Qualitative SIM			Quantitative SIM
	Water	Sewerage	Billing	Contacts
Internal sewer flooding	✘	✓	✘	✓
Mean zonal compliance	✓	✘	✘	✓
Category 3 pollution incidents	✘	✓	✘	✓
Water supply interruptions	✓	✘	✘	✓
Water quality contacts	✓	✘	✘	✓

Source: Economic Insight Analysis

6.2.2. Quantitative analysis

The existence of data on firms' SIM performance and the common performance commitments means that it is possible to conduct some limited quantitative analysis to inform our consideration of the complementarity of the SIM with the common performance commitments. This analysis is, however, limited by available data.

As the structure of the SIM changed after PR14, to maintain consistency and maximise the number of data points, we have examined correlation between the SIM from PR09 and the performance commitments over this period. For the three water-related performance commitments, we therefore have sufficient observations to provide a robust analysis.

To examine all five performance commitments, however, we are left with three years of data across ten WaSCs, giving 30 observations in total. With five common performance commitments, and possible additional controls, there are at most 25 degrees of freedom. This means that the robustness of this analysis is inherently questionable. This caveat should be borne in mind throughout this analysis.

6.2.2.1. Correlations

Examining data on the relationship between SIM scores and the water-related performance commitments for all companies does not suggest any material correlation. This can be seen in the scatter diagrams below, which show no obvious pattern. Regression analysis confirms this, with

none of the water commitments being statistically significant in explaining the SIM as a whole and its qualitative and quantitative constituents (see section **Error! Reference source not found.** below).

Figure 29: SIM scores and mean zonal compliance
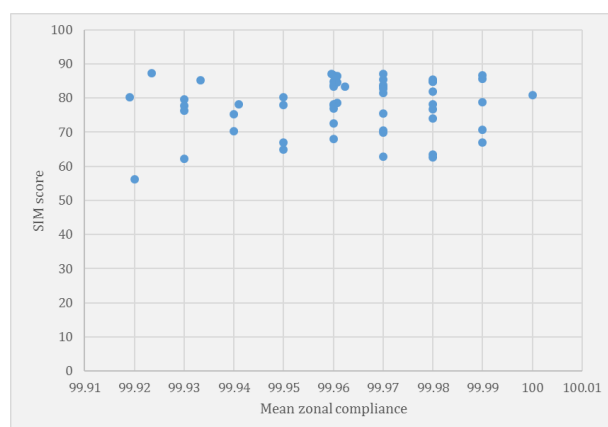
Source: Economic Insight Analysis of Ofwat PR14 comparative assessment data (2011/12 – 2013/14)

Figure 30: SIM scores and water quality contacts

Source: Economic Insight Analysis of Ofwat PR14 comparative assessment data (2011/12 – 2013/14)

Figure 31: SIM scores and water supply interruptions
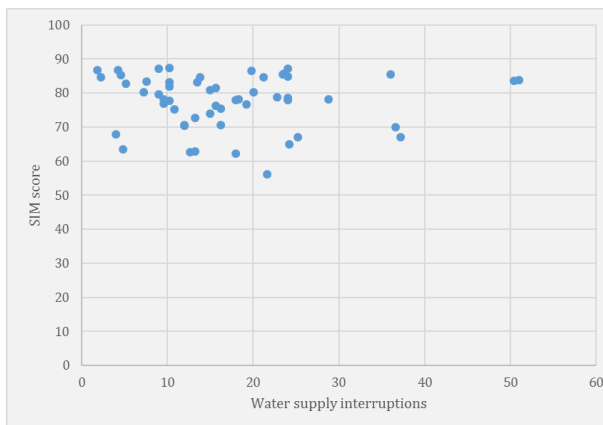
Source: Economic Insight Analysis of Ofwat PR14 comparative assessment data (2011/12 – 2013/14)

6.2.2.2. Regression analysis

When we used regression analysis to examine the correlation between SIM scores and the performance commitments for WaSCs only, the sewerage commitments and mean zonal compliance appeared to have statistically significant relationships with qualitative SIM scores, and the R-squared of the regressions increased from around 2 per cent to over 40 per cent, meaning that the inclusion of these variables meant that the regressions explained more of the variation in SIM scores. When we tested the robustness of these results by including additional controls, and by expressing the variables in logarithms or with respect to firms' performance relative to the frontier, sewer flooding and mean zonal compliance

remained statistically significant in most but not all specifications.

Table 9: Statistical significance of performance commitments, qualitative SIM

Controls	Sewer flooding	Mean zonal compliance	Category 3 pollution incidents
None	✓	✓	✓
Time trend	✓	✓	✓
Logs	✓	✓	✘
Time trend & logs	✓	✓	✘
Relative performance	✘	✘	✓
Relative, trend and logs	✓	✓	✓

Source: Economic Insight Analysis of Ofwat common performance commitment data

There may be some evidence of overlap with internal sewer flooding, but we would need more data points to confirm such a relationship. Further, even if this relationship could be substantiated, as we set out above, this does not translate into certain conclusions about the complementarity of the existing SIM.

If consumers do give systematically lower SIM scores when internal sewer flooding incidents are more common, this could mean that firms have two reasons to avoid such incidents – first to avoid falling behind on performance on the commitment, and secondly to avoid reductions in their SIM score. One could think of the relationship with SIM scores as reflecting strong consumer feelings about this performance commitment, in which case this effect appears appropriate. When setting the balance of incentives, however, if such a relationship were substantiated with more data, then this potential relationship should be considered.

6.2.3. Detailed results

The tables below set out the detailed results of our econometrics analysis of SIM scores and the other common performance commitments from data used for rewards and penalties at PR14 (using data from 2011/12 to 2013/14).

Table 10: Results of analysis of all companies

Dependent variable	SIM		Qualitative SIM		Quantitative SIM	
Companies	All		All		All	
Logs	No		No		No	
Relative performance	No		No		No	
Other controls	None		None		None	
R-squared	0.0212		0.0187		0.0440	
Observations	54		54		54	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
Constant	-5328.2413	0.3650	1044.6210	0.5655	-6372.8624	0.1730
Internal sewer flooding	-	-	-	-	-	-
Mean zonal compliance	54.0884	0.3581	-10.0204	0.5815	64.1088	0.1707
Cat 3 pollution incidents	-	-	-	-	-	-
Water supply interruptions	-0.0323	0.7723	-0.0221	0.5236	-0.0102	0.9079
Water quality contacts	-0.1450	0.8550	0.0026	0.9917	-0.1475	0.8142

Source: Economic Insight Analysis of Ofwat common performance commitment data

Table 11: Results of analysis of WaSCs only

Dependent variable	SIM		Qualitative SIM		Quantitative SIM	
Companies	WaSCs		WaSCs		WaSCs	
Logs	No		No		No	
Relative performance	No		No		No	
Other controls	None		None		None	
R-squared	0.3856		0.4417		0.3289	
Observations	30		30		30	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
Constant	7338.2754	0.3552	5999.8180	0.0398	1338.4574	0.8111
Internal sewer flooding	-55.7794	0.0211	-24.1545	0.0060	-31.6249	0.0609
Mean zonal compliance	-72.4305	0.3615	-59.5179	0.0413	-12.9126	0.8177
Cat 3 pollution incidents	-0.0288	0.0098	-0.0095	0.0150	-0.0192	0.0144
Water supply interruptions	-0.0632	0.6126	-0.0211	0.6340	-0.0421	0.6350
Water quality contacts	-1.2133	0.2579	-0.4337	0.2541	-0.7796	0.3057

Source: Economic Insight Analysis of Ofwat common performance commitment data

Table 12: Results of analysis of WaSCs, variables in logs

Dependent variable	SIM		Qualitative SIM		Quantitative SIM	
Companies	WaSCs		WaSCs		WaSCs	
Logs	Yes		Yes		Yes	
Relative performance	No		No		No	
Other controls	None		None		None	
R-squared	0.4359		0.4765		0.3668	
Observations	30		30		30	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
Constant	445.9221	0.3386	672.5238	0.0311	99.2111	0.8987
Internal sewer flooding	-0.1458	0.0047	-0.1093	0.0013	-0.1926	0.0225
Mean zonal compliance	-95.8757	0.3434	-145.2414	0.0319	-20.7250	0.9026
Cat 3 pollution incidents	-0.0435	0.1623	-0.0195	0.3243	-0.0718	0.1714
Water supply interruptions	-0.0575	0.1185	-0.0331	0.1601	-0.0919	0.1384
Water quality contacts	0.0322	0.3527	0.0218	0.3276	0.0506	0.3863

Source: Economic Insight Analysis of Ofwat common performance commitment data

Table 13: Results of analysis for WaSCs, time trend

Dependent variable	SIM		Qualitative SIM		Quantitative SIM	
Companies	WaSCs		WaSCs		WaSCs	
Logs	No		No		No	
Relative performance	No		No		No	
Other controls	Time trend		Time trend		Time trend	
R-squared	0.5174		0.5254		0.4698	
Observations	30		30		30	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
Constant	5794.9879	0.4210	5542.6207	0.0446	252.3672	0.9606
Internal sewer flooding	-40.7715	0.0684	-19.7084	0.0197	-21.0631	0.1796
Mean zonal compliance	-57.1110	0.4278	-54.9795	0.0463	-2.1315	0.9667
Cat 3 pollution incidents	-0.0282	0.0059	-0.0093	0.0120	-0.0188	0.0093
Water supply interruptions	0.0530	0.6648	0.0134	0.7670	0.0396	0.6498
Water quality contacts	-1.4127	0.1510	-0.4928	0.1735	-0.9199	0.1885
Time trend	3.6416	0.0197	1.0788	0.0559	2.5628	0.0213

Source: Economic Insight Analysis of Ofwat common performance commitment data

Table 14: Results of analysis for WaSCs, variables in logs, time trend

Dependent variable	SIM		Qualitative SIM		Quantitative SIM	
Companies	WaSCs		WaSCs		WaSCs	
Logs	Yes		Yes		Yes	
Relative performance	No		No		No	
Other controls	Time trend		Time trend		Time trend	
R-squared	0.5234		0.5282		0.4586	
Observations	30		30		30	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
Constant	484.3487	0.2708	692.2388	0.0234	161.9490	0.8262
Internal sewer flooding	-0.1124	0.0249	-0.0922	0.0069	-0.1382	0.0958
Mean zonal compliance	-104.2382	0.2751	-149.5318	0.0240	-34.3782	0.8300
Cat 3 pollution incidents	-0.0522	0.0811	-0.0240	0.2192	-0.0861	0.0899
Water supply interruptions	-0.0174	0.6568	-0.0125	0.6302	-0.0264	0.6910
Water quality contacts	0.0193	0.5589	0.0152	0.4884	0.0295	0.5978
Time trend	0.0427	0.0515	0.0219	0.1260	0.0697	0.0604

Source: Economic Insight Analysis of Ofwat common performance commitment data

Table 15: Results of analysis for WaSCs, independent variables relative to frontier performance

Dependent variable	SIM		Qualitative SIM		Quantitative SIM	
Companies	WaSCs		WaSCs		WaSCs	
Logs	No		No		No	
Relative performance	Yes		Yes		Yes	
Other controls	None		None		None	
R-squared	0.315074		0.352638		0.277613	
Observations	30		30		30	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
Constant	83.2902	0.0000	44.3491	0.0000	38.9410	0.0000
Internal sewer flooding	2.3265	0.1472	1.1190	0.0579	1.2075	0.2772
Mean zonal compliance	-0.8494	0.5647	-0.8604	0.1150	0.0110	0.9915
Cat 3 pollution incidents	4.1666	0.0117	1.3770	0.0199	2.7896	0.0152
Water supply interruptions	-0.2197	0.8913	0.0032	0.9957	-0.2229	0.8429
Water quality contacts	1.9807	0.2305	0.6919	0.2463	1.2889	0.2635

Source: Economic Insight Analysis of Ofwat common performance commitment data

Table 16: Results of analysis for WaSCs, independent variables relative to frontier performance, variables in logs, time trend

Dependent variable	SIM		Qualitative SIM		Quantitative SIM	
Companies	WaSCs		WaSCs		WaSCs	
Logs	Yes		Yes		Yes	
Relative performance	Yes		Yes		Yes	
Other controls	Time trend		Time trend		Time trend	
R-squared	0.5260		0.5203		0.4809	
Observations	30		30		30	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
Constant	74.2664	0.0000	41.3567	0.0000	32.9097	0.0000
Internal sewer flooding	2.5963	0.0618	1.2085	0.0227	1.3879	0.1544
Mean zonal compliance	-1.1827	0.3503	-0.9709	0.0477	-0.2118	0.8127
Cat 3 pollution incidents	4.0379	0.0049	1.3343	0.0115	2.7036	0.0076
Water supply interruptions	-0.5455	0.6910	-0.1049	0.8380	-0.4406	0.6525
Water quality contacts	2.0164	0.1541	0.7037	0.1825	1.3127	0.1912
Time trend	4.3444	0.0040	1.4407	0.0094	2.9037	0.0064

Source: Economic Insight Analysis of Ofwat common performance commitment data

6.3. Technical details on incentives under monopoly

This section presents further technical details on the underlying economics of service provision under monopoly supply. It first discusses incentives for customer service under unregulated and regulated monopoly supply, before going on to discuss how incentives can be used to address under-provision.

6.3.1. Service incentives under monopoly supply

Monopoly supply is associated with market failures that result in inefficient outcomes. Whereas in competitive markets, firms generally have incentives to provide the outcomes that customers value – provided that consumers value them sufficiently highly relative to the costs of providing them.

Regulatory interventions such as the SIM are generally thought of as providing incentives that promote more efficient outcomes, closer to those that would be observed in a competitive market. For instance, Ofwat's statement of

principles for PR14 referred to the purpose of economic regulation as being to replicate the effects of competitive pressure.[18] The principles for Water 2020 also include explicit reference to the use of incentives to promote allocative, dynamic and productive efficiency.[19]

An efficient set of outcomes would involve consumers paying a price and receiving bundles that were *efficient*. Specifically, they would be **allocatively efficient, technically efficient and dynamically efficient**.

In the absence of regulation, monopoly suppliers have incentives both to overcharge for what they supply, and to provide customer service outcomes that are below efficient levels. Specifically, they will only provide customer service levels up to the point at which the marginal revenue associated with such levels equals their marginal cost.

In reality, firms are subject to price caps on the amount that they can charge for these bundles. Unlike in a competitive market they receive no potential benefit through higher revenues if they provide better outcomes or service levels.

[18] *'The role and design of incentives for regulating monopoly water and sewerage services in England and Wales – a discussion paper.' Ofwat (2010); page 9*

[19] *'Towards Water 2020 – meeting the challenges for water and wastewater services in England and Wales: discussion document.' Ofwat (2015); page 42.*

Moreover, unlike an unregulated monopolist, there is no marginal revenue associated with improving their outputs. In this case, price-capped firms have an incentive to minimise their costs.

These situations are shown in the figure below. Under cost minimisation provision is even lower than under monopoly, which again is lower than the efficient competitive level.

**Figure 32: The monopoly problem for service levels**
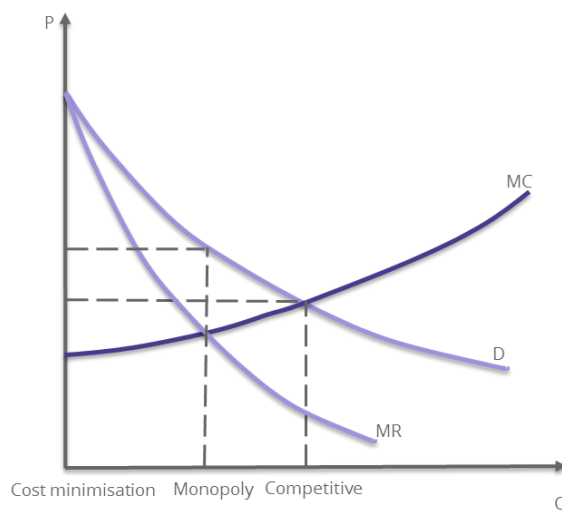
Source: Economic Insight analysis

This leads to outcomes and service levels that are inefficient, specifically:

» Outcomes and service levels are underprovided, such that marginal benefit of providing each outcome or service level is in excess of the marginal cost of doing so.

» There is below optimal investment in projects that would provide new outcomes or increase service levels.

The precise extent of under-provision, and therefore the strength of incentive required, depends on the balance of costs and the value that consumers place on particular outcomes and service levels. As shown in **Error! Reference source not found.** below, under-provision will be greater for outcomes that are more valued and have lowest costs.

Figure 33: Relationship between value, cost and under-provision
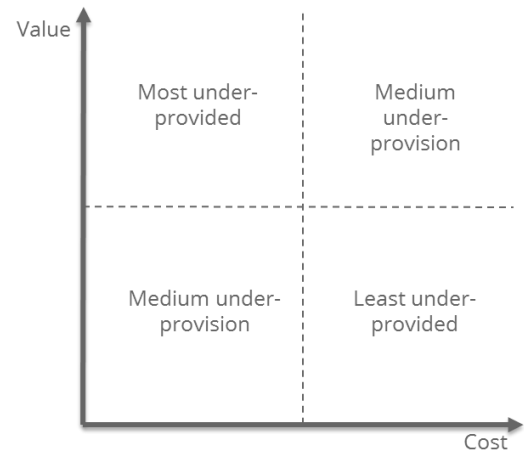
Source: Economic Insight analysis

6.3.2. Using incentives to address under-provision

To address the issue of under-provision and induce firms to produce outcomes that they would not otherwise, regulators often provide incentives to firms to increase their provision. In doing this, they need to ensure that the incentives that they provide mean that the marginal benefit to the firm (marginal revenue) from producing an output or increasing service levels equals or exceeds the marginal cost of this to the firm. If less than marginal cost, firms will not increase their provision.

If the marginal benefit to the firm from the incentive mechanism equals the marginal cost to the firm, then the firm is essentially indifferent between producing the output and not producing it. In this case the incentive would exactly mimic outcomes in a competitive market and the outcome would be efficient. All of the benefits from increasing efficiency would accrue to consumers. In this case the 'amount' of incentive is determined by the costs incurred by the firm in providing an efficient outcome level.

A regulator could target other efficient outcomes, however, up to all of the additional benefits accruing to firms. The 'amount' of incentive required in this case would equal the total value of benefits from increased efficiency.

This is shown in the figure below, comparing the competitive situation relative to a monopoly that does not face a price cap. If firms are provided

with incentives that induce 'competitive' outcomes, then total incentives would equal the grey square. Other incentives consistent with efficient outcomes go up to the situation in which incentives equal total additional benefits from achieving efficiency, in which case firms would receive the grey square, plus the purple triangle.

Figure 34: Addressing monopoly issues with incentives
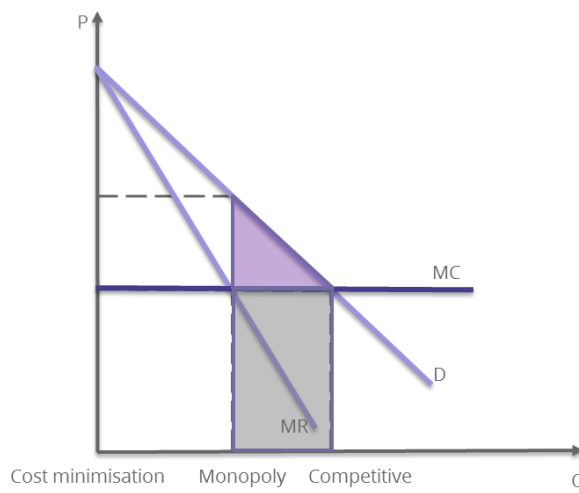
Source: Economic Insight analysis

In practice regulators are confronted with an information problem when trying to address this issue. It is impossible to know precisely what are the outcomes that consumers value most, the costs that firms incur to deliver them, and therefore the extent to which they are underprovided. This means that it is not feasible for regulators to target precisely what would happen in a competitive market. (Indeed, if it were, regulators could simply make rules that required companies to produce particular outcomes or service levels, rather than constructing incentive schemes.)

Instead, regulators generally provide incentives that encourage firms to deliver outcomes that would otherwise underprovided, without targeting an exact level of provision as representing the exact competitive target level. Rather than aiming for firms to be indifferent between producing and not producing an outcome, as would occur in a competitive market, the benefits from increasing efficiency are generally shared between firms and consumers. Our cross-sectoral review suggests

that there is a particular focus in the utilities sector on basing outcomes and incentive levels on consumers' willingness to pay.

In such circumstances, the size of the incentive is generally thought of as being consistent with what consumers would be willing to pay for the outcome, and splits the benefits of improved efficiency between firms and consumers, in a way that ensures that benefits accrue for both groups. In other words, expectations of consumers' willingness to pay inform the total amount of funding that is available to provide incentives for outputs.

6.4. Outcomes frameworks and water company performance commitments

This section summarises outcomes frameworks used in other sectors and service-related water company performance commitments.

6.4.1. Outcomes frameworks

We present our detailed analysis of outcomes frameworks used by 14 regulators and public sector bodies and provide summary details of the frameworks that we examined.

6.4.1.1. The role and purpose of outcomes frameworks.

In general, we found that outcomes frameworks for private companies that are subject to economic regulation generally seek both to monitor companies' performance and provide incentives to achieve particular outcomes. Outcomes frameworks involving public bodies tend only to involve reputational incentives and monitoring. They are also often used to identify and publicise best practice.

6.4.1.2. The outcomes that were identified for inclusion.

There is wide variation in the nature of the outcomes that are included within the frameworks. The number of outcomes included ranges from one to 20, with other frameworks including even more sub-measures. While many frameworks retain a focus on the same set of outcomes over time, some frameworks (particularly those associated with health) change the outcomes they focus on. Such frameworks are generally used for monitoring

purposes, rather than for applying financial incentives.

6.4.1.3. How outcomes were chosen

There is generally a strong focus in the frameworks for private utility companies on choosing outcomes (and associated incentive levels) based on consumer preferences and willingness to pay. This focus appears to be absent from other frameworks.

Across the frameworks, the extent of observability and measurability appears to have been a key driver of the outcomes that were selected for inclusion, with a lack of readily available and observable measures being one reason for focusing on more high-level outputs. As this report makes clear, however, one might wish to incentivise high-level outcomes for other reasons, for instance to provide firms with flexibility to provide the most valued outcomes.

As we set out in this report, an important consideration in choosing outcomes is the extent to which they were in company control. This is reflected in some of the frameworks that we examined. For example, Ofcom has undertaken a range of work to understand how variations in both cost and outcome performance may or may not be within Royal Mail's control.

6.4.1.4. How incentives were set

As discussed above, private companies subject to economic regulation are often set direct financial incentives, whereas public bodies tend to be subject to reputational incentives or monitoring. Where financial incentives are set, targets for firms' performance are often selected by benchmarking performance using either historical data or by comparing performance against other firms in the same sector or in related sectors.

6.4.1.5. Information on individual frameworks

We provide more information on the outcomes frameworks that we analysed in the table below. We provide additional information on regulators' approaches to measuring customer service in the final section.

Table 17: Summary information on outcomes frameworks

National Health Service (NHS) England	
Outcomes Framework	The NHS Outcomes Framework 2015/16.
Role and purpose	The Outcomes Framework is designed to provide a national overview of how well the NHS is performing. It is the primary accountability mechanism between the Secretary of State for Health and NHS England. It improves quality throughout the NHS by encouraging a change in culture and behaviour focused on health outcomes not process.
Outcomes identified for inclusion	The outcomes framework contains five domains which are derived from the mandate. The mandate represents a key document published by the Department of Health which sets out the important objectives for NHS England.
	The five key domains are associated with a total of 68 indicators which are used to measure performance in health and care systems at the national level. They are not an exhaustive list but are designed to provide a view of the current state of services. The five domains which encompasses the 68 indicators are: » Preventing people from dying prematurely; » Enhancing quality of life for people with long-term conditions; » Helping people to recover from episodes of ill health or following injury; » Ensuring that people have a positive experience of care; and » Treating and caring for people in a safe environment and protecting them from avoidable harm.
How measures were translated into concrete incentives	Success is measured against the average level of improvement, progress in reducing health inequalities and unjustified variation. The Secretary of State publishes an annual assessment of progress.

Public Health England (PHE)	
Outcomes Framework	Public Health Outcomes Framework (PHOF).
Role and purpose	The framework has been designed to refocus attention on achieving positive health outcomes for the population and reducing inequalities in health and to shift away from focusing on process targets and performance managing local areas.
Outcomes identified for inclusion	PHOF have identified two outcomes that align with their vision. The outcomes are: (i) increased healthy life expectancy; (ii) reduced differences in life expectancy and healthy life expectancy between communities The two outcomes cover both morbidity and mortality, addressing within-area differences and between-area differences.

Public Health England (PHE)	
Measures	The PHOF consists of 68 public health indicators. This includes, two overarching indicators or outcomes mentioned above and 66 more focused indicators which are grouped into four 'domains': » improving the wider determinants of health; » health improvement; » health protection; » healthcare public health and preventing premature mortality. The domains were designed to provide performance measures against the two key outcomes. This is because the outcomes themselves may take years if not decades to see any significant change so the indicators were developed to provide information on progress made annually at the national and local level. Some indicators have sub-indicators. In total there are 147 indicators and sub-indicators.
How measures were translated into concrete incentives	PHOF is accompanied by an interactive web tool which allows local authorities to assess their progress relative to national and regional averages. It further enables them to set their progress based on a benchmark set by others. From 2014/15 the Department of Health is running a pilot Health Premium Incentive Scheme. The scheme rewards local authorities when they are able to make improvements in line with selected indicators from the PHOF. For 2015-16 payment to local authorities is based on "successful completion of drugs treatment" with combined PHOF data for opiate and non-opiate users; and an indicator selected by local authorities (or the default smoking prevalence in over 18s).

Northern Ireland Authority for Utility Regulation (NIAUR)	
Outcomes Framework	Water and sewerage services price control 2015-21.
Role and purpose	The Utility regulator's role is to protect the interest of consumers in relation to the supply of water and the provision of sewerage services. Its primary duties are to: » Protect the interests of consumers; » Ensure that NI Water carries out its functions properly; and » Ensure that NI Water is able to finance its functions.
Outcomes identified for inclusion	The final determination outlines four outcomes: » Lower bills for most water and sewerage customers » Improved efficiency » Continued significant investment in water and sewerage services » Improvements in levels of service

Northern Ireland Authority for Utility Regulation (NIAUR)	
Measures	There are 28 measures in total for water and 17 measures for sewerage services. These are grouped into five categories: customer service, quality water/sewerage, water/ sewerage outputs, serviceability and new output measures.
	The Consumer Engagement Oversight Group worked collaboratively to develop a plan for consumer engagement and extensive research was conducted to understand the aspects of water and sewerage services which matter most to households and businesses. The research conducted in 2013, involved both qualitative research and quantitative research. The qualitative research consisted of 12 focus group discussions with 97 domestic consumers and 17 in-depth interviews conducted on a one-to-one basis with non-domestic consumers. The quantitative research comprised of a survey of 1,032 randomly selected households across NI and 512 telephone surveys with non-domestic consumers.
How measures were translated into concrete incentives	NI Water is required to prepare a Monitoring Plan for PC15 outlining the outputs and milestones for delivery. NIAUR will publish Cost and Performance Report annually setting out the progress the company has made in delivering the targets in the Monitoring Plan. The report will set out NIAUR's view on sustainable efficiencies that could be made.

Water Industry Commission for Scotland (WICS)	
Outcomes Framework	Strategic Review of Charges 2015-2.
Role and purpose	The Framework serves to help monitor performance and also to help incentivise Scottish Water to deliver the required outcomes, both through reputational and financial mechanisms. Outcomes were largely determined by the Government, as set out in the Ministerial Directions. The Water Commission established a Customer Forum, to identify customer priorities and to secure the best outcome for customers. The Forum engaged in direct discussions with Scottish Water over several months, representing the concerns of customers and leading to a better outcome for customers.
Outcomes identified for inclusion	There are eight outcome areas: drinking water quality; environment; nature; waste; climate change adaption; climate change mitigation; flood risk management; and security of supply.
Measures	The Customer Forum and Scottish Water have three customer service measures and four service improvement measures. The customer service measures are: household customer experience measure; the non-household customer experience measure; and the high esteem test. The service improvement measures are in relation to external sewer flooding, visible leakage, reduction in carbon emissions, managing responses to extreme weather events, and minimising the number of complaints that are escalated up to or upheld by the Scottish Public Services Ombudsman.

Water Industry Commission for Scotland (WICS)	
How measures were translated into concrete incentives	WICS uses a combination of financial and reputational incentives. The household customer experience measure is built from a quantitative and qualitative component similar to SIM by Ofwat. Similar measure is also developed for non-household customers so that comparison can be made with other water companies.

Commission for Energy Regulation (CER) – Irish Water	
Outcomes Framework	Irish Water (CER Interim Price Control 2014 to 2016).
Role and purpose	The overarching Irish Water goal is to deliver safe, affordable and environmentally compliant water services to all customers.
Outcomes identified for inclusion	The key priorities are identified under the following categories: drinking water quality and capacity, waste water quality and capacity, water conservation, asset renewal, replacement and improvement, metering and asset data.
Measures	The CER has not set targets for any of the outcomes.
How measures were selected	The CER drew on the OPA from other jurisdictions as a basis for the outputs framework that the CER would monitor Irish Water on. While the CER is not intending to implement a full OPA framework in Ireland, the metrics behind the OPA were used as a starting block for the CER's research. The CER also reviewed metrics which are in place for energy utilities in Ireland
How measures were translated into concrete incentives	Not applicable.

Transport Scotland	
Outcomes Framework	NTS Framework – January 2016.
Role and purpose	The 2006 NTS ensures that transport acts as an enabler of economic growth – to support businesses in achieving their local, national and international objectives and to improve the lives of individuals and communities by connecting them with their economic future. It is aimed at providing sub-policy areas and sub-national levels with a framework to refer to when they develop and implement their own plans / policies / strategies and monitor and report on their delivery.
Outcomes identified for inclusion	There are three Key Strategic Outcomes: (i) improved journey times and connections, to tackle congestion and lack of integration and connections in transport; (ii) reduced emissions, to tackle climate change, air quality, health improvement; and (iii) improved quality, accessibility and affordability, to give choice of public transport, better quality services and value for money, or alternative to car.

Transport Scotland	
Measures	The achievement of three transport-specific National Indicators is tracked by the National Performance Framework through Scotland Performs: (i) reduce traffic congestion; (ii) increase the proportion of journeys to work made by public or active transport; and (iii) reduce death on Scotland's roads.
How measures were translated into concrete incentives	Not applicable.

Welsh Government	
Outcomes Framework	The Wales Transport Strategy 2008.
Role and purpose	The goal of the Wales Transport Strategy is to ensure a future that promotes sustainable transport networks, provide service based on social justice and equality, while strengthening Wales' economic and social life. The Welsh Government has a key role in delivering the Welsh Assembly Government's strategic agenda.
Outcomes identified for inclusion	The transport strategy identifies 17 outcomes in total, they are grouped into three high - level outcomes related to: social, economic and environmental outcomes.
Measures	Each of the 17 outcomes is accompanied by one or more indicators which are used as a guideline as to whether the outcomes are met. There are 35 measures in total.
How measures were translated into concrete incentives	Reputational incentives only.

Office of Road and Rail (ORR)	
Outcomes Framework	Network Rail's outputs and funding 2014-19.
Role and purpose	Network Rail is accountable to the ORR for the day to day running of its service. This includes monitoring train punctuality, reliability of its service and upkeep of rail assets. The framework consists of outputs which Network Rail must deliver for the money it receives. It also includes indicators which are used for monitoring purposes. A failure to deliver outputs could potentially be considered a licence breach, however, not meeting an indicator measure will not breach the licence rather it is used to detect trends about likely breach of the output.
Outcomes identified for inclusion	The ORR set outputs in the areas that matter most to passengers, freight customers and the industry. The ORR has set 58 outputs and given passenger operators and Network Rail the flexibility to agree further annual outputs for punctuality and cancellations.

Office of Road and Rail (ORR)	
Measures	The key measures are grouped into eight categories: train service reliability, enhancements, health and safety, network availability, network capability, stations and depots, asset management and environment.
How measures were translated into concrete incentives	Both the Secretary of State and Scottish Ministers provide funding, and include requirements to be included as outputs. However, mostly reputational incentives apply.

Transport for London	
Outcomes Framework	Monitoring under the Mayor's Transport Strategy (MTS).
Role and purpose	The MTS sets out the transport vision for London and is built around six transport goals. These long-term outcomes sought by MTS are monitored through the collection and publication of a set of 24 Strategic Outcome Indicators (SOIs). These are used to evaluate, over the longer-term, the overall direction of progress in relation to Mayoral transport goals, primarily so as to provide appropriate feedback to the on-going policymaking process.
Outcomes identified for inclusion	MTS is built around six transport goals: (i) Supporting economic development and population growth; (ii) enhancing the quality of life for all Londoners; (iii) improving the safety and security of all Londoners; (iv) providing opportunities for all Londoners; (v) reducing the contribution of transport to climate change and improving its resilience to the impacts of climate change; and (vi) supporting the delivery of the London 2012 Olympic and Paralympic Games and their legacy.
Measures	The long-term transport outcomes sought by the MTS are monitored through a set of 24 quantitative SOIs.

These indicators are 'outcome-based', reflecting changes in conditions experienced by Londoners. They provide a manageable means of assessing the overall direction and pace of change in relation to MTS goals. They do not cover all aspects of transport that will be of interest and do not, of themselves, provide a detailed understanding of topical transport issues.

The SOIs quantify progress in the delivery of the MTS in order to facilitate a broad understanding of the 'totality of effects' of the strategy's interventions on transport and wider quality of life in London. They consulted with key stakeholders across Surface Transport, including Strategy and Planning policy leads and sponsors across Surface Transport. |
| **How measures were translated into concrete incentives** | Not applicable |

New Zealand Transport Agency (NZTA)	
Outcomes Framework	NZTA Statement of Intent 2013-16.
Role and purpose	NZTA's purpose is to deliver transport solutions for New Zealand (NZ) on behalf of the Government. Outcomes framework provides a holistic picture of the NZTA's service delivery and investment performance.
Outcomes identified for inclusion	NZTA Outcomes Framework has four key outcomes which describe their outlook for the transport sector in the future. These outcomes are associated with their long-term goals which in turn delineate the organisation's direction and action over the next 20 years. The four key objectives are: (i) planning the land transport network; (ii) providing access to and use of the land transport system; (iii) managing the state highway network; and (iv) investing in land transport.
Measures	Measures are set based on outcome and goals. The three areas they were set on are: (i) value for money; (ii) service quality; and (iii) customer satisfaction. Targets were set for all of the years in the SOI period (here 2012-13 to 2015-16).
How measures were translated into concrete incentives	Monitoring framework whereby NZTA reports against the above outputs in order to demonstrate how it performs for the funding it receives.

Ofgem	
Outcomes Framework	T1 - electricity and gas transmission.
Role and purpose	RIIO-T1 is designed to encourage network companies to: (i) put stakeholders at the heart of their decision-making process; (ii) invest efficiently to ensure continued safe and reliable services; (iii) innovate to reduce network costs for current and future consumers; and (iv) play a full role in delivering a low carbon economy and wider environmental objectives
Outcomes identified for inclusion	RIIO-T1 set out six primary outputs: (i) safety; (ii) reliability; (iii) availability; (iv) customer satisfaction; (v) connections / wider works (new investment); and (vi) environmental outputs.
Measures	There are 13 measures of the six outcomes. Additionally, there are five Network Output Measures which are indicators of asset performance. Where possible, the target was informed by economic value of a unit of output to consumers. In addition, some of the outputs are linked with licence conditions and statutory obligations, where the targets are already set within these (in some instances set by other organisations).

How measures were translated into concrete incentives	A number of the incentives are linked to the percentage of allowed revenue. To maintain strong output incentives and appropriate revenue allowances for specific activities, the caps and collars need to adjust in response to changes in revenue to better reflect the true change in network total expenditure (totex) and other in-period adjustments over the price control period.

Ofcom	
Outcomes Framework	Annual monitoring update on the postal market: 2015
Role and purpose	Ofcom monitors the postal market to secure the provision of a universal postal service such that it is financially sustainable and efficient.
Outcomes identified for inclusion	Ofcom monitors Royal Mail's Quality of Service very closely, specifically through various product-related and service-oriented targets. It also uses consumer and business surveys to gauge satisfaction with the postal sector. Royal Mail is monitored in four areas: » Financial performance » Operating performance » Impact on customers and consumers » Impact on competition
Measures	Royal Mails performance is measured against 22 key metrics. First Class and Second Class national targets were set below 100% to allow for commonly occurring disruptions that may arise in the transportation, processing and delivery of mail. There are 121 Post Code Areas (PCAs) in total. Royal Mail is required to deliver 91.5% of all First Class single piece mail the day after collection to 118 PCAs. This target is set to ensure that Royal Mail provides a good level of service across the UK –not just in more densely populated areas. Additionally, they also take into account exceptional circumstances.
How measures were translated into concrete incentives	If Royal Mail fails to meet its Quality of Service targets, Ofcom will consider opening a formal investigation. Ofcom believes that the risk of investigation and enforcement action incentivises Royal Mail to meets its Quality of Service targets for universal services. Many of the current targets were developed from Royal Mail's own internal targets and have been in place since 2001. Historical performance suggests that all of the targets are achievable.

Department for Education (DfE)	
Outcomes Framework	Secondary School Accountability System 2013

Department for Education (DfE)	
Role and purpose	The objective of the framework is to make schools accountable for their performance through the reputational implications of having key performance metrics published and available to stakeholders. It is also aimed at highlighting the progress that students make while at school or college.
Outcomes identified for inclusion	Headline outcomes published to give a snapshot of the performance of schools include: (i) progress; (ii) attainment; (iii) English and Maths GCSEs; iv) retention; and v) destinations.
Measures	The five outcome areas have five associated measures. Two headline measures have associated 'minimum standards', which are in effect targets to be achieved. Progress is determined by the number of grades (or proportion of one grade) below the national average level of progress of zero. Combined completion and attainment measure = number of grades (or proportion of grade) below the national average.
How measures were translated into concrete incentives	The performance measures about each school are distributed in three ways: (i) on the school website, (ii) on school performance tables and (iii) in a data portal which will provide a single point of access to include all information on schools and pupils. Enforcement action can be taken should a school's performance fall below the specified minimum standard. If intervention is necessary, or if an institution is judged inadequate by Ofsted, they will be required to improve. They will come under additional scrutiny, until the DfE considers that they have sufficiently improved. However, should a provider come under scrutiny and not improve, then this could result in the withdrawal of funding, closure, changes in leadership or a school becoming a sponsored Academy.

Source: Economic Insight research

6.5. Customer service measurement

This section provides more detail on innovative measures used in competitive markets to measure customer service performance, and measures used by water companies.

6.5.1. History of customer service measurement

Formal customer satisfaction research developed in the 1960s and 70s. In many mature markets it was, by this time, becoming harder for companies to differentiate themselves from their competitors on the basis of their products alone. Companies instead began to look to their service offer. The paper-based surveys of this era were expensive to administer and so at first customer satisfaction surveys were typically only run by large institutions and companies.

Initially research focused on customers' previous usage patterns as a way of measuring their loyalty to a brand. However, it became evident that past behaviour was in fact a poor predictor of future behaviour. The first of the three main measures of customer experience, customer satisfaction, emerged during the late 1970s and early 1980s. It is typically asked of people who have had a recent interaction with a company and asks customers to rate their satisfaction, often on a five-point scale.

R. L. Oliver published two influential papers in 1977 and 1980 outlining his Expectations Confirmation Theory.[20, 21] The theory held that the level of customer satisfaction could be understood by measuring the gap between expectation and delivery. Various research models such as SERVQUAL were developed using this theory and customer satisfaction emerged as the most widely used question to encapsulate consumers' overall experience during this period.

During the 1990s and early 2000s questions began to be raised as to how well customer satisfaction was correlated with customers' future purchasing behaviour. For example, in the Harvard Business Review in 1995 Thomas Jones and Earl Sasser Jr published research to show that customers who were fairly satisfied with a service were not that much more likely to repurchase than those who were actually dissatisfied.[22] They did, however, find that customers who had very high levels of satisfaction (who they identified as being 'truly loyal') were much more likely to remain loyal to the brand. They also noted that customer satisfaction levels varied significantly across markets and that customer satisfaction did not, by itself, provide sufficient diagnostic information to drive company strategy and innovation.

In 1996 Fred Reichheld published his book 'The Loyalty Effect' which continued Jones' and Sasser's criticism of customer satisfaction.[23] He pointed out that 60-80% of customers that switched providers were in fact satisfied or very satisfied with the product or service.

We note that these issues do not apply in the context of the new incentive mechanism. The purpose of measurement is not to provide diagnostic information, while correlation with customer loyalty is of less interest in the context of a monopoly provider.

In 2003 Reichheld went on to develop net promoter score, which measures the number of strong advocates of the brand (likely to recommend to a friend) minus the number of detractors as measured on an 11-point scale. It has been widely used by many companies as a way of predicting customer loyalty and is well known and understood in the private sector.

In recent years there has been something of a backlash against net promoter score. Criticisms levelled at the measure include the fact that, like customer satisfaction, net promoter score provides no diagnostic information to understand why customers are giving the scores that they do and that it ignores the views of a substantial proportion of customers who fall in the middle of the scale and who are therefore neither advocates nor detractors of the brand.

The correlation between net promoter score and future customer behaviour has also been questioned by some and like, all three of the main customer experience measures, there are concerns that it only provides a narrow understanding of the complex issue of customer satisfaction. Net promoter score was adopted by many companies as a company indicator and there have been problems with some erratic movements in the measure.

The most recent of the three main measures of customer satisfaction is Customer Effort Scoring. This was introduced in 2012 in a Harvard Business Review article by Max Dixon in which he argued that, in many situations, it is unnecessary to delight customers when what they primarily wish for is for their interactions with companies to be easy and straightforward. This in many ways runs counter to the criticisms posed to customer satisfaction in the 1990s.

In 2013 Moira Clark of Henley Business School and Andrew Bryan of BT published a study of early adopter companies of customer effort score. They found that companies believed it to be a useful additional measure to net promoter

---

[20] *'Effect of Expectation and Disconfirmation on Post Exposure Product Evaluations – an Alternative Interpretation.'* R.L. Oliver, Journal of Applied Psychology (1977).

[21] *'A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions.'* R.L. Oliver, Journal of Marketing Research (1980).

[22] *'Why Satisfied Customers Defect.'* T. Jones & W. Earl Sasser Jr., Harvard Business Review (1995).

[23] *'The Loyalty Effect.'* F. Reichheld (1996).

(and sometimes customer satisfaction) because of its versatility and its correlation with loyalty.

Furthermore, the measure was readily understood by staff, though in some cases it required a change of mind set to move from the desired goal of 'delighting' a customer. Many companies have adapted the question to suit their particular contexts. For example, BT created a 'net easy' score based on the question "How easy was it to get the help you wanted today?"

In parallel with the evolution of question wording, there has also been a gradual change in interviewing method. During the 1990s postal and in person interviews gave way to computer-assisted telephone interviewing (CATI) and then in the 2000s to online and mobile interviewing techniques. These new methods allow for timelier, cheaper and less intrusive interviewing. Companies are also beginning to look to a broader range of inputs to assess customer satisfaction, such as through social media listening and through the wealth of personal information now possible using big data. However, it is currently difficult to condense inputs from these new methods into the single

number approach to customer satisfaction measurement that many organisations desire.

Throughout the evolution of customer satisfaction monitoring, the three key areas of debate have been

» which measure is most likely to act as a predictor of future behaviour;
» whether it is possible for any one measure to encapsulate all the complexity of the relationship with each customer and provide sufficient direction for company development and improvement; and
» the extent to which different measures are appropriate and useful for different industries.

6.5.2. Innovative ways of measuring satisfaction

We examined some innovative approaches to measuring satisfaction.

» Rant & Rave
» Forrester Customer Experience Index
» YouGov BrandIndex
» Brain Juicer tracking
» Research communities and panels.

Table 18: Innovative approaches to customer service measurement

Rant & Rave	
Description	Rant & Rave is a customer service measurement platform.[24] Customers who have recently interacted with an organisation are contacted via text/email to answer a simple indicator such as customer effort, net promoter score or satisfaction. They are then invited to leave verbatim comments which are coded using sentiment analysis.
Advantages	» Rapid feedback and likely to generate good sample sizes. » Simple and easily understandable methodology. » Can provide contact details of those giving low scores to allow company to try to further resolve issue. » Already used by several Water Companies. » Works best as a 'full circle' approach of measurement and customer feedback leading to direct action.
Disadvantages	» Little benefit in using simply to track key performance indicators. » Limitations of sentiment analysis. » Can only ask 1 or 2 indicators. » Provides no feedback from general customer base.

---

[24] rantandrave.com

Rant & Rave	
Conclusion	Rant & Rave is a platform, rather than an experience measure per se.

Forrester Customer Experience Index	
Description	The Forrester Customer Experience Index is a research 'system' to measure customer experience.[25] It calculates two measures: customer experience quality and customer experience loyalty. Customer quality: » Effectiveness (the experience was effective in delivering value to customers). » Ease (for customers to get value from the experience). » Emotion (customers feel good about their experience). Customer loyalty: » Retention (likelihood that a customer will keep their existing level of business). » Enrichment (likelihood that a customer will purchase additional products and services). » Advocacy (likelihood that a customer will recommend you to others). The survey then measures a number of industry-specific diagnostic measures to provide understanding as to core drivers.
Advantages	» Comprehensive measurement of customer experience and understanding of key drivers. » Data available for benchmarking.
Disadvantages	» Not an efficient way of measuring a limited number of indicators. » Customer loyalty measures inappropriate in a monopoly situation.
Conclusion	This Index focuses on a much wider set of indicators than is necessary for the new incentive mechanism, so would not be an efficient way to measure outcomes for the new incentive mechanism.

YouGov BrandIndex	
Description	Uses the nationally representative YouGov panel to measure public perceptions of brand measures.[26] Measures a variety of brand metrics including 'buzz' (positive or negative mentions in the media), customer satisfaction, purchase intent and employee advocacy. Companies can also pay to add a limited number of additional questions.
Advantages	» Benchmarking data available for other companies/industries buying survey. » Able to provide quick responses.
Disadvantages	» General sample – respondents would be asked about water companies not in their area. » Some questions are irrelevant for our purposes (e.g. purchase intent).

[25] *go.forrester.com/data/cx-index*

[26] *brandindex.com*

YouGov BrandIndex	
Conclusion	The focus here is on brand perception rather than customer satisfaction, so would not be suitable for the new incentive mechanism.

Brain Juicer Tracking	
Description	Brain Juicer use findings from the field of behavioural economics to provide a new spin on traditional brand tracking.[27] Behavioural economics suggests that it is false to believe that our behaviour can be accurately recalled and predicted and that we are often unconscious of our true reasons for our behaviours. Brain Juicer's tracker focusses on: » Fame – how readily your brand comes to mind. » Feeling – how strongly people feel about your brand. » Fluency – how easily people recognise your brand.
Advantages	» There is some acceptance that respondents are less able to accurately record their behaviour, feelings and decision-making processes than has been previously thought. » This approach provides a very different way of measuring brand opinion and is already being used by companies in competitive markets.
Disadvantages	» Does not measure the core indicators of satisfaction. » Likely to be controversial due to the novelty of its approach, which is based on brand opinion rather than service opinion. » Difficult to use in a monopoly environment.
Conclusion	The focus here is on brand perception rather than customer satisfaction, so would not be suitable for the new incentive mechanism.

Research Communities and Panels	
Description	Research communities are a panel of respondents who have been recruited to participate for a period of time in a number of research studies online. They are typically used by large organisations who need regular access to research sample. Communities can become involved in discussions online or can respond to polls and surveys. They can also become involved in co-creation in developing new products and services. Companies can also access panels run by third-party agencies in order to reduce costs, for example Wessex Water's use of the Future Focus panel.
Advantages	» Cost-effective choice for companies needing to regularly undertake research. » Provides rich information.
Disadvantages	» Not suitable for measuring indicators over time. » Community members can become highly involved with the brand and thus unrepresentative of the wider customer base. » Not cost-effective for limited research.

---

[27] brainjuicer.com

Research Communities and Panels	
Conclusion	These methods are unsuitable for the future incentive mechanism because they focus on the views of a fixed set of customers over a period of time. This does not appear appropriate for looking at water company customer service issues, which are likely to occur intermittently.

Source: Caroline Thompson Associates research

6.5.3. Measures used by water companies

We have examined publically available information on the measures used by a sample of water companies. This is not a comprehensive analysis of companies' approaches, but is intended to give examples of the approaches that firms use in practice. We note that the purpose of these approaches is different from the purpose of the incentive mechanism. For firms, it will also be important to gather diagnostic information.

This suggests that many conduct surveys both among a general sample of their customer base and those who have had a specific interaction. Those who have had recent contact with the company are better able to provide detailed feedback on the service they received enabling the water companies to diagnose where problems are occurring.

Some methodologies (e.g. Rant and Rave) allow faster feedback than is possible currently via SIM. Again, this helps water companies to react more quickly when problems occur.

With respect of customer satisfaction, there is a mix of usage of customer satisfaction, net promoter score and customer effort score with no one measure being consistently used across the companies. Some companies look for additional feedback on company image e.g. perceptions of value for money.

Overall, there is no one research model that is overwhelmingly favoured by companies in the samples. However, smaller groups of companies are buying some 'off-the-shelf' methodologies: UKCSI, Rant and Rave and Future Focus Customer Panel approaches.

In general, therefore, the measures that water companies use are either the standard measures such as customer satisfaction, or suffer from the same problems as the innovative measures that we examined.

Table 19: Examples of water company measures

Water Company	Surveys
Affinity	'Rant and Rave' survey. Customers in contact with Affinity are contacted by text or email within 24 hours and invited to participate in a survey. First, they are asked an NPS measure and are then invited to comment on reasons for their score. Affinity subsequently re-contacted customers providing an NPS score of six or lower to try to resolve any outstanding problems. Affinity has used this service from February 2011 and are reported to receive 3000 pieces or customer feedback a month with a 15% response rate. Over 50% of those completing the survey leave a verbatim comment which is then analysed using Rant and Rave's sentiment analysis software.
Dee Valley Water	Rant & Rave. Dee Valley asks customers to rate them on a scale of 1 to 5 rather than using NPS.

Water Company	Surveys
Dŵr Cymru	Rant and Rave, UKCSI. Biannual CATI survey to measure trust. Customers are asked if they 'trust Welsh Water to do the right thing.' NPS asked via online survey (sample unclear).
Scottish Water	Rant & Rave.
South East Water	Continuous survey among a general customer sample to assess customer satisfaction with a number of specific areas e.g. "appearance, taste and odour of water", "Interruptions to supply".
Sutton and East Surrey Water	Regular survey of general customer base to measure customer satisfaction, value for money and affordability.
Thames Water	Rant & Rave. Measures customer satisfaction, customer effort and NPS.
Wessex Water	Customer Satisfaction among sample of people who have experienced a recent interaction. Image tracking among a general sample Customer Panel run by Future Focus Research asks about a different topic (e.g. home water check service) every 3 months. Receive approximately 1200 replies. Future Focus also list Bristol Water and Welsh Water among their clients.
Yorkshire Water	UKCSI. SIM Replica Survey – replicates SIM survey among customers who have had a recent interaction, but asks 200 customers per month (SIM conducted quarterly). Customer Voice – Text survey to customers at different stages of customer journey to ask customer satisfaction (five-point scale). Domestic Tracker – Survey among a general sample of 300 customers to ask customer satisfaction and value for money.

Source: Caroline Thompson Associates research

6.5.4. Measures used by other regulators

We have analysed the customer service-related measures that other regulators use to monitor and set incentives.

Table 20: Summary of other regulators' approaches to customer service monitoring and incentives

Body	Outcomes framework	Purpose	Surveyed satisfaction	Other relevant outcomes (e.g. value for money)
WICS	Strategic Review of Charges 2015-21	Monitoring	Called customer experience New measure, similar to SIM, has a quantitative and qualitative component but the sample includes views of customers who have not made direct contact	High esteem test: used to compare Scottish Water's reputation with the UK utilities, along with other companies and brands across all sectors
Ofgem	RIIO-ED1 price control	Financial incentives	Broad measure of customer satisfaction includes surveyed satisfaction	Also includes a 'complaints metric' based on the number of complaints, with a weighting similar to the quantitative SIM. Companies are also assessed with respect to the quality of their stakeholder engagement
ORR (Rail)	Network Rail's outputs and funding 2014-19	Monitoring	National Passenger Survey	-
TfL	Monitoring under the Mayor's Transport Strategy	Monitoring	Surveyed travel experience	-
NZ Transport Agency	NZ Transport Agency Statement of Intent 2013-16	Monitoring	Online feedback survey and CATI	Operational targets e.g. % of activities delivered on time
Ofcom	Annual monitoring update on the postal market	Monitoring	Two separate surveys to track use of and attitudes to post	-
Ofcom	The Customer Experience	Monitoring	Ofcom monitors overall satisfaction and complaint numbers for fixed line, mobile, fixed broadband, digital TV and bundled product providers	-

Body	Outcomes framework	Purpose	Surveyed satisfaction	Other relevant outcomes (e.g. value for money)
NIAUR	Northern Ireland Water	Monitoring	Call handling satisfaction score (1-5)	Percentage: billing contact dealt with within five working days, written complaints dealt with within ten working days, metered customers received bill based on a meter reading, calls not abandoned, calls not receiving the engaged tone

Source: Economic Insight research

Further Information Please Contact

Christopher Pickard

e: christopher.pickard@economic-insight.com
t: +44 (0) 207 100 3746
m: +44 (0) 7572 711802

Economic Insight Limited

125 Old Broad Street
London
EC2N 1AR
www.economic-insight.com