



Europe Economics

The introduction of B-MeX into the Business Retail Market

Final report to Ofwat

31 March 2021

Europe Economics
5 Chancery Lane
London EC4A 1BL

Tel: (+44) (0) 20 7831 4717
Fax: (+44) (0) 20 7831 4515

www.europe-economics.com

Europe Economics is registered in England No. 3477100. Registered offices at 5 Chancery Lane, London, EC4A 1BL.

Whilst every effort has been made to ensure the accuracy of the information/material contained in this report, Europe Economics assumes no responsibility for and gives no guarantees, undertakings or warranties concerning the accuracy, completeness or up to date nature of the information/analysis provided in the report and does not accept any liability whatsoever arising from any errors or omissions.

© Europe Economics. All rights reserved. Except for the quotation of short passages for the purpose of criticism or review, no part may be used or reproduced without permission.

Contents

Executive Summary.....	1
The Case for a B-MeX Incentive.....	1
Key Policy Issues for B-MeX Incentive	2
Recommendations and Next Steps	4
1 Introduction.....	5
1.1 Objective of the study	5
1.2 Approach to the study.....	5
1.3 Structure of the report	6
2 Could a B-MeX Incentive Lead to Improved Outcomes for Customers?	7
2.1 Overview of the business retail market	7
2.2 Activities wholesalers and retailers undertake in the business retail market	8
2.3 Influences on wholesaler engagement with end customers	10
3 Is It Practical to Design a B-MeX Incentive?.....	13
3.1 Lessons learnt from the sector and other sectors	13
3.2 Basic design of a B-MeX incentive	14
4 Weighting of Customers	17
4.1 Introduction	17
4.2 Description and analysis of issues.....	17
4.3 Recommendations	24
5 Use of Complaints	25
5.1 Introduction	25
5.2 Description and analysis of issues.....	25
5.3 Recommendations	32
6 Relative or Absolute Performance.....	35
6.1 Introduction	35
6.2 Description and analysis of issues.....	35
6.3 Recommendations	40
7 Regulatory Instrument for B-MeX Incentive	41
7.1 Introduction	41
7.2 Description and analysis of issues.....	41
7.3 Recommendations	46
8 Size of Financial Incentive.....	48
8.1 Introduction	48
8.2 Description and analysis of the issue	48
8.3 Recommendations	53
Appendix I: Lessons Learnt from the Water and Other Sectors	54

Scottish business retail market..... 54

Rail 55

Electricity 56

Aviation 58

Executive Summary

Europe Economics was commissioned by Ofwat to explore the concept and feasibility of a business customer measure of experience (B-MeX). The B-MeX incentive would focus on the services provided by wholesalers to non-household customers in the business retail market with a view to incentivising wholesalers to improve their service offering. In completing this project we worked with a steering group consisting of members from Ofwat, Consumer Council for Water (CCW) and MOSL. Whilst these parties provided input into the project, the analysis and recommendations contained in this final report reflect the views of Europe Economics. They do not necessarily reflect the official views of Ofwat, CCW or MOSL.

The research comprised two phases:

- Phase 1 investigated whether it was **possible and practicable to design a B-MeX incentive** that reflected wholesaler performance in terms of the services provided to end customers, that were not significantly influenced by retailer performance. This included considering how wholesaler and retailer activities may influence the outcomes experienced by end customers, as well as identifying any gaps in current regulatory arrangements that may be addressed through a B-MeX incentive.
- Phase 2 considered **five key policy issues relevant for the design of a B-MeX incentive**. This included considerations around weighting the experience of different customers, the use of complaint data, relative or absolute measures of performance, the regulatory instrument to be used for the scheme, and the size of any financial incentives attached.

We summarise each of these phases of the research below, and then draw out our overall recommendations and suggested next steps from the study.

The Case for a B-MeX Incentive

In the business retail market wholesalers and retailers are responsible for providing different services to end customers. For example, wholesalers' activities typically relate to water and wastewater operations, involving the maintenance of assets, responding to service failures (water pressure, leakage, sewage, interruptions), providing services around connections, meter installation and repair. At the same time retailers focus more on customer-facing activities such as billing, performing meter readings, and acting as the first line of communication between end customers and wholesalers.

In terms of the current regulatory measures in place to influence wholesalers' conduct in the business retail market and/or affect the experience of non-household customers, we identify two important gaps in current industry arrangements that might warrant a B-MeX incentive.

1. First, the potential gaps in the current industry framework mainly relate to qualitative aspects of wholesaler-led end-customer interactions. For example, a business customer's experience interacting with a wholesaler to have its meter repaired will be influenced by a number of qualitative aspects including:
 - its experience making the appointment (e.g. how quickly it could arrange an appointment and whether convenient time slots are available);
 - its experience of having a contractor on site to physically repair the meter (e.g. whether workers are polite and show due consideration); and
 - the degree to which the wholesaler completes the job in an effective and efficient way (e.g. the customer is not passed from pillar to post and the job is genuinely completed in a single visit).

How well the wholesaler delivers on such factors typically involve judgement and they are not easily quantifiable but customers know when they have experienced good or bad service.

2. Second, the current Market Performance Framework (MPF) standards relate to market and operational processes, while the retailer measure of experience (R-MeX) measure focuses on qualitative aspects of wholesaler-retailer interactions. This means that the current MPF lacks the links between the services provided by wholesalers and end customer experience.

Taken together, this suggests that a measure focusing on the qualitative aspects of interactions between wholesalers and end business customers appears to be missing from the suite of incentives in place through current industry arrangements. Many of these qualitative aspects would be difficult to quantify objectively, yet in other market settings would be something that firms subject to competitive pressures would be keen to do well for fear of losing customers to their rivals. A B-MeX regime could potentially provide wholesalers with a similar incentive to seek to improve end customer experience. To work, it would require direct input from business customers, in the form of survey evidence, complaints data, or a mix of the two.

We suggest surveys of business customer experience used to generate a B-MeX score should focus on those business customers who have directly interacted or communicated with their wholesaler recently.¹ The general population of business customers may not be able to distinguish between the role of retailers and wholesalers. Moreover, the focus of a B-MeX regime is to improve wholesalers' dealings with business customers so it will be more practical to select a sample that includes business customers with recent experience of the activities that are the focus of the B-MeX incentive.

The scope of a B-MeX measure should extend to all activities where the wholesaler has a material role to play in ensuring good end-customer outcomes. In many cases both the retailer and wholesaler have a role to play in realising such outcomes. At first glance, it may seem unfair to penalise or reward a wholesaler for something that is not entirely their responsibility. However, we caution against adopting a B-MeX regime that only provides incentives for services where no responsibility could be attached to the retailer. If there are some activities where wholesalers' behaviour is partly, but not fully, responsible for the customers' level of satisfaction, B-MeX should be incentivising the wholesaler to attempt to improve end-customers outcomes, by directly improving their service offering or engaging with the retailer with a view to realising better outcomes for business customers. It would be wrong to send a message that there is no need to do anything because the retailer also has a role to play.

Key Policy Issues for B-MeX Incentive

We have considered five policy issues that are important in designing a B-MeX incentives. In assessing them, we have considered whether they are likely to serve the customer interest, be effective in creating the right incentives for wholesalers, and the cost and practical implications of each issue. While we have provided recommendations, there may be scope for further consideration of the issues as part of the more detailed design work.

- **Weighting of customers:** The design of a B-MeX incentive would need to appropriately consider whether the experience of certain customers should be given more weight than others. For example, assigning more importance to responses of the largest business customers could mimic the incentives present in competitive markets on firms to focus service provision on those who consume the most. It is also a way of giving more weight to customers who are considered more informed, perhaps better able to distinguish between the services of the wholesaler from the service of the retailer. A survey limited to customers that have recently engaged with wholesalers is already likely to have a sample that is made up of a larger share of large businesses than the general population of businesses. We recommend not

¹ This does not, however, rule out including indirect interactions between business customers and wholesalers where the influence of the wholesaler on end customer experience is material.

making any further adjustments, other than possibly some statistical representation weightings to ensure that the responses received align with the population from which the sample was drawn (i.e. the businesses that recently interacted with a wholesaler).

- **Use of complaints data:** There is the question of what use to make of information on complaints for the purposes of a B-MeX incentive. First, complaints data could help identify the issues that are of concern to customers, and their relative importance. Second, they could be used as a component of the B-MeX scores, such that wholesalers have reputational and possibly financial incentives to minimise the number of complaints. Overall, we recommend using complaints data to help with the design of a B-MeX regime, but that complaints should not be used in the actual scheme. Complaints data are a valuable source of information, identifying instances where the customer detriment has been sufficiently large to prompt a customer to make a complaint. Furthermore, they have the advantage of capturing any and all instances where a business customer is unhappy with service, whereas surveys may limit the range of activities that a business customer can comment on. Nevertheless, we do not recommend using the complaints data to determine the B-MeX score. There are practical problems, including the fact that there are likely to be relatively few written complaints against the smaller water-only companies. There are also potential incentive problems, including the potential for parties to try and game the system, although there may be steps that could be taken to address risks of gaming.
- **Relative or absolute measures of performance:** We have considered whether a B-MeX scheme should be based on a relative measure of performance meaning a company's performance is evaluated relative to that of other companies or an absolute measure of performance where company performance is assessed against absolute benchmarks. Relative measures of performance may create similar incentives to those that firms operating in a competitive market might face, encouraging wholesalers to compete and do better than other wholesalers on the business customer experience they provide. This also means that relative measures of performance will entail implicit targets that evolve over time to reflect improvements in what the industry is able to achieve. Wholesalers will permanently have incentives to strive to outdo other wholesalers. In addition, the informational requirements for regulators in the case of relative measures are less, as regulators do not need to set a target and determine what is achievable. The major drawback with a relative performance regime is the possibility that there are factors outside the control of wholesalers that are difficult to control for when making comparisons between wholesalers. One candidate would be the large differences in the market shares of different retailers in different regions. We recommend that the default should be for the B-MeX incentive to be based on a relative measure of performance. The threshold for accepting deviations from this and adopting an absolute measure of performance should be high. If there are some activities where an absolute measure is deemed appropriate, we would still recommend retaining a relative measure of performance for other activities to be covered by the B-MeX incentive.
- **Regulatory instrument:** Different regulatory instruments may be used for implementing and administering a financial B-MeX incentive including using the MPF, the price control framework, or offering direct financial compensation to customers. While the latter possibility has the attraction of compensating those business customers directly affected by poor wholesaler service, the need to rely on customer feedback to determine qualitative aspects of the service provided means that such an approach is unlikely to be practical. In principle, requiring wholesalers to pay penalties into the central fund of the MPF or introducing additional Outcome Delivery Incentives (ODIs) in the price-control framework could work to incentivise improved wholesaler performance through a financial B-MeX incentive.
- **Financial versus reputational incentives:** The design of a financial B-MeX scheme will also need to consider the size of any financial incentive attached to ensure that wholesalers are incentivised to provide better service. The size of any financial incentive would also need to consider any other reputational and financial incentives faced by wholesalers through existing industry arrangements. A balance needs to be struck so as not to encourage companies to put a disproportionate emphasis on qualitative aspects of the business customer experience at the expense of core operational requirements. More generally, it should be set

having regard to customers' willingness to pay for an improved level of service. Further work would be needed in this regard. In the immediate future, it is likely that a B-MeX incentive should only have reputational consequences.

Recommendations and Next Steps

Overall, we think that a B-MeX incentive could fill important gaps in the existing regulatory framework governing water (and wastewater) companies. The focus of such a scheme should be on the qualitative aspects of services provided by wholesalers to end business customers. We suggest that the B-MeX incentive should primarily depend on evidence from surveying business customers who have had recent dealings with wholesalers. The precise design of the survey questions and the specific wholesaler interactions covered by the incentive are beyond the scope of this work.

In terms of next steps for the development and implementation of B-MeX, we recommend that a working group with input from key market participants (e.g. wholesalers, retailers, CCW, MOSL and Ofwat) should seek to develop a pilot B-MeX incentive measure that could, in principle, run from April 2022. This could be informed by CCW complaints data and any evidence that wholesalers and retailers have already collected that might be relevant. This working group will also need to consider how a B-MeX incentive should be funded (for example could it use penalties paid by wholesalers for failing to comply with existing MPF standards?)

Early on, we recommend there should be no financial rewards or penalties attached to this B-MeX measure. The case for publishing the scores (and therefore creating reputational incentives) during the pilot phase is more finely balanced. We favour publication unless there are compelling reasons to believe that the pilot B-MeX measure is unlikely to be reliable (e.g. if there are concerns that the data from different regions will not be compiled on a like-for-like basis).

By the time of the next price review (2024), the expectation should be that there will be a financial B-MeX incentive regime (unless evidence gathered leads Ofwat to conclude that financial incentives are inappropriate). This should hold true whether Ofwat ultimately decides to include the B-MeX incentive – either with or without end customer compensation – within PR24 or Ofwat or industry opts to incorporate it in the MPF. The industry should start collecting the additional evidence needed to refine the B-MeX incentive measure and determine appropriate financial incentives having regard to this timeline. Lessons learnt from the pilot regime should also be used.

1 Introduction

Europe Economics has been commissioned by Ofwat to explore the concept and feasibility of a business customer measure of experience (B-MeX) focusing on the services provided by wholesalers to non-household customers² in the business retail market. In completing this project, we worked with a steering group consisting of members from Ofwat, CCW and MOSL. Whilst these parties provided input into the project, the analysis and recommendations contained in this final report reflect the views of Europe Economics. They do not necessarily reflect the official views of Ofwat, CCW or MOSL.

1.1 Objective of the study

The study looks at the overall concept of the B-MeX incentive as well as considering some of the high-level design issues associated with a B-MeX scheme. It is a preliminary exploration of the issues, and is not intended to consider more detailed design or implementation issues associated with a B-MeX incentive scheme.

The first objective of the project is to provide a recommendation around whether it is possible and practical to design a B-MeX incentive that reflects wholesaler performance in terms of the services provided to end customers, that were not significantly influenced by retailer performance.

The study then explores some of the high-level design issues associated with a B-MeX incentive (including the weighting of customers, use of complaints, relative or absolute measures of performance, the regulatory instrument and the question of financial and/or reputational incentives). We provide recommendations on these issues where there are clear economic arguments supporting a design decision. In other cases, the report considers the relative merits and disadvantages of different options for these issues, and indicates the further analysis and information required before a decision is made.

1.2 Approach to the study

Our approach to the requirements has relied primarily on desk-based research, supported by engagement with relevant stakeholders.

For the desk-based research we used materials published by water companies to explore the services offered both to retailers and end business customers, and reviewed the relevant documentation published by Ofwat and MOSL to identify the incentives companies face through both the price control framework and the MPF. In addition, we also considered the information published by CCW on the complaints received that could be attributed to wholesalers. Finally, we analysed relevant documents published by regulators to identify any lessons to be learnt from the water sector and other regulated sectors.

We spoke with four wholesalers and some members of the UK Water Retailer Council, as well as with Ofwat, MOSL and the CCW to gather directly information and opinions from stakeholders on matters where the desk-based research was insufficient. We have also had meetings with a project steering group consisting of staff from Ofwat, MOSL and CCW.

The analysis and assessment of policy issues were based on our internal experience and expertise in incentive design issues, and through looking at the relevant economic and regulatory literature.

² The terms “non-household customer”, “business customer” and “end customer” are used interchangeably in this report.

1.3 Structure of the report

This report first addresses the high-level questions about whether a B-MeX regime might have merit and how practical it would be to introduce such a scheme. Later chapters then look at some of the important design issues that need to be resolved prior to implementing a B-MeX incentive.

Chapter 2 provides an overview of the business retail market, followed by a summary of the activities typically undertaken by wholesalers and retailers in the market. It identifies some potential gaps in terms of interactions between wholesalers and end-customers where the incentives for providing a good service may be poor given current industry arrangements. The chapter considers what the activities may be where a B-MeX regime might fill a gap and provide incentives to wholesalers to improve customer outcomes.

Chapter 3 summarises some of the lessons from other sectors and considers some of the questions around how customers might be surveyed and the scope of activities that may be covered in a B-MeX regime. The chapter also identifies the key policy issues that warrant further consideration and the assessment criteria we have used when evaluating different policy options.

Chapter 4 considers the options around weighting customers in terms of its merits in improving incentives on wholesalers to focus on the end customer experience in tailoring the provision of wholesale services.

Chapter 5 explores two broad ways in which complaints data might be used for B-MeX. First, it investigates whether complaints data could help in identifying the issues that are of concern to customers, and their relative importance. Second, it explores whether complaints information could be used as a component of the B-MeX scores, such that wholesalers have reputational and possibly financial incentives to minimise the number of complaints.

Chapter 6 considers the arguments for and against the use of absolute and relative measures of performance in the context of the B-MeX incentive.

Chapter 7 explores three different regulatory instruments that may be used to implement a financial B-MeX incentive: the use of a central fund (as per the current MPF); the price control framework; and offering direct compensation to customers.

Chapter 8 looks at the options and issues around the size of any financial incentives for a B-MeX scheme.

The Appendix summarises the lessons that may be learnt from the water and other regulated sector in terms of the design of quality and service incentives and customers measures of satisfaction. In particular, it looks at the Scottish business retail market, as well as the rail, electricity and aviation sectors.

2 Could a B-MeX Incentive Lead to Improved Outcomes for Customers?

2.1 Overview of the business retail market

A water retail market in England (the business retail market) was opened to competition in April 2017.

The opening of the business retail market brought significant changes to the way in which non-household customers buy their water services. Eligible non-household customers in England are now free to choose their retailer and are no longer restricted to buying these services from the regional monopoly provider.³ Instead, business customers can shop around and negotiate better deals if they are not satisfied with the services they get from their current provider of retail services. Prior to the opening up of the business retail market, non-household customers dealt directly with the retail arm of the incumbent regional monopoly provider. For a business with operations in more than one part of the country, the retail services would be provided by many different wholesalers; now such a business has the opportunity to employ the same retailer across all the regions it operates in.

There is significant diversity amongst these non-household customers in terms of size, consumption and expenditure on water and wastewater services. Table 1 below provides an overview of business customers based on these characteristics.

Table 1: Summary characteristics of business water customers in 2019-20

Customer size (number of employees)	Eligible businesses as % of total	Consumption (%)	Range of expenditure (£)	Average annual expenditure (£)
Micro (0-9)	86%	11%	<£500 - £9,999	£350
SME (10-249)	13%	25%	<£500 - >£100,000	£2,500
Small (10-49)	11%	n/a	<£500 - £100,000	n/a
Medium (50-249)	2%	n/a	<£500 - >£100,000	n/a
Large (250+)	<1%	64%	£1,000 - >£100,000	£35,000

Source: Ofwat (2020) "State of the market 2019-20: review of the third year of the business retail water market" [[online](#)].

The business retail market for non-household customers in Wales operates differently to the English business retail market. While in England business customers are able to switch between retailers, under the Water Act 2003 in Wales only non-household customers with annual consumption exceeding 50 million litres of water can select their water retailer.⁴ Business customers in Wales cannot choose their wastewater retailer.

2.1.1 Business retail market structure

At the time of market opening, the structure of the business retail market was shaped by various forces.

³ Broadly speaking, eligible customers that are able to switch in the business retail market are (i) business customers supplied by an appointed company whose area is wholly or mainly in England; and (ii) business customers supplied by an appointed company whose area is wholly or mainly in Wales and using a minimum of 50 mega litres of water a year. For further details on eligible customers, please refer to Ofwat [[online](#)].

⁴ Senedd Research (2018) "The Water Industry in Wales" [[online](#)].

- A number of companies already active in Scotland (which opened its retail business market in 2008⁵) acquired business customers from the wholesalers who exited the business retail market. For example, Business Stream first started operating in Scotland and then also extended its operations to England, acquiring the non-household customer base of Southern Water in 2017 and of Yorkshire Water Business Services and Three-Sixty in 2019.⁶
- Some wholesalers de-merged their business retail operations to financially and legally separate Retailer undertakings, which took on the business customer book of the associated wholesalers. Hence a number of retailers directly associated with certain wholesalers became the main retail provider in a region. For example, Water Plus is a joint venture between Severn Trent Water and United Utilities while Wave is a joint venture between Anglian Water Business and NWG Business.
- Smaller, independent retailers such as Everflow entered the market. Such retailers did not purchase a business customer book from a wholesaler so their market share is entirely dependent on organic growth through customer acquisition.

The opening of the market also allowed for some non-household consumers to become self-suppliers. Instead of using the services of a retailer to interact with the wholesaler(s), these customers deal directly with the wholesaler regarding water provision related issues, and therefore act as their own retailer. Such an arrangement is typically favoured by large companies. As of March 2020, there were 12 self-suppliers active in the market.⁷

In 2020 there were over 20 retailers active in the English business retail market, and 15 retailers in the Welsh business retail market, with some overlap between the retailers operating in these markets. They provide services to approximately 1.2 million non-household customers, including businesses, charities and public sector organisations.

Looking at the distribution of Supply Point IDs (SPIDs) in 2019-20,⁸ the retailers with the highest market shares are Water Plus (30 per cent), Castle Water (22 per cent), Wave (16 per cent) and Business Stream (15 per cent). The remaining retailers account for approximately 16 per cent of the market combined. These numbers mask the fact that market shares are very concentrated within given wholesaler regions. The market shares of in-area retailers (i.e. those who acquired the customer base of the previous monopoly incumbents), have been slowly declining, with customers moving to new-entrants or, in a few cases, switching to self-supply. Nevertheless, the market shares of the in-area retailers remain large. In 2017-18, 99.5 per cent of the market belonged to in-area retailers, 0.4 per cent to new entrants and 0.1 per cent to self-supply. These shares had changed to 96.8 per cent, 2.7 per cent and 0.5 per cent in 2019-20, respectively.⁹

2.2 Activities wholesalers and retailers undertake in the business retail market

The opening of the business retail market also implied a separation in the provision of different services to non-household customers. This section explores the type of activities that are provided by wholesalers and retailers in the business retail market, and considers the boundaries between the services provided by wholesalers and retailers.

⁵ Business Stream (2018): “Scotland marks decade of water competition” [\[online\]](#).

⁶ Business Stream (2019): “Business Stream buys non-domestic customer base of Yorkshire Water Business Services and Three-Sixty” [\[online\]](#).

⁷ Ofwat (2020): “State of the market 2019-20: review of the third year of the business retail water market” [\[online\]](#).

⁸ SPID is an identifier of the water and waste-water connection to the company’s premise. Each company will have either one or two SPIDs, depending if they require both water and waste-water connection or only one of the two. SPIDs are different for each premise.

⁹ Ofwat (2020) “State of the market 2019-20: review of the third year of the business retail water market” [\[online\]](#).

Activities typically undertaken by wholesalers in the business retail market relate to waste and wastewater operations, involving the maintenance of assets, responding to service failures (water pressure, leakage, sewage, interruptions), providing services around connections, meter installation and repair, and trade effluents.¹⁰

In performing these activities wholesalers in general may interact with end customers in two different ways:

- Some of the operational and maintenance activities involve site visits to the business property of end customers where **wholesaler interacts directly with end customers**. This would typically involve physical services such as installing a meter or replacing a broken one, fixing a leaking pipe, responding to incidents involving low water pressure, or sewer flooding. In these cases, the behaviour of wholesaler representatives during the visit (e.g. whether they have taken their boots off where appropriate, how they communicated with the business owner or representative), the way in which the appointment was set up and carried out (e.g. how quickly an appointment could be arranged and whether convenient time slots are available etc.), the degree to which the wholesaler completes the job in an effective and efficient way (e.g. the customer is not passed from pillar to post and the job is genuinely completed in a single visit), and the consequences or follow up to the visit (e.g. whether and how the wholesaler communicated the old and new meter reads to the customer and retailer in the case the wholesaler replaced a meter) could have a significant impact on end customer experience and their business. In the case of direct interactions between wholesalers and end customers the responsibility around the qualitative aspects of service provision to end customers during site visits rest directly with wholesalers.
- Other activities may involve direct (or in some cases indirect) **communication between wholesalers and end customers**, for example around leakage allowance policies or other requests raised by businesses. In these cases, it may not always be clear whether the wholesaler or retailer is responsible for drawing up and communicating the policy to the end customers. For example, in the case of leakage allowances, wholesalers would typically set the policy around any refund however it is generally the retailer's responsibility to communicate the policy as well as any reasons for a rejected application to the business customers. Initiatives around informing and educating customers around leakage policies and similar issues could be seen as falling under both wholesalers' and retailers' responsibilities.

The services provided to individual end customers may also vary. Some of the variation will reflect the fact that different customers will have different requirements for interacting the wholesaler (e.g. in case of service interruptions requiring urgent action from the wholesaler, a business may prefer to contact the wholesaler directly rather than raise the issue first with their retailer). In other cases, there may be regional differences in how wholesalers deal with non-household customers. For example, leakage allowance policies are not uniform across the country. This in turn means that a business located in a wholesaler region with a less stringent leakage allowance policy may get its leakage allowance approved while a customer located in a wholesaler region with a more stringent allowance is likely to see its application rejected, even if the circumstances under which the leakage occurred are similar. Consequently, a consumer in a region with less stringent allowances may judge that the wholesaler has done a 'good job' in terms of explaining the policy and dealing with their application which is less likely to be the case for a customer whose application was rejected.

By contrast, retailers focus more on customer-facing activities, which typically involve billing (and any related issues), performing meter readings, and acting as the first line of communication between end customers and wholesalers when business customers want to raise any queries, issues or complaints regarding the services delivered by wholesalers. In the latter case, where wholesalers are notified of any issues around the service (e.g. a supply interruption or a sewer flooding event), retailers' level of involvement in resolving these issues can also vary. For example, one water company we spoke with noted that some retailers are keen to be

¹⁰ For example, for the range of services offered by Thames Water see: Thames Water (2020): "Wholesale Service Offering" [[online](#)].

involved in communications between the wholesaler and end-customer, while others let wholesalers interact directly with end customers with a view to resolving the issue in a timely manner.

Given the variation across wholesalers in terms of the services directly provided to end business customers and possible differences between the activities performed by both wholesalers and retailers across different regions, the boundaries between wholesaler and retailer activities may not always be clear-cut to end customers. For example, customer research provided by one wholesaler indicates that customers are often not fully aware of the distinction between wholesalers and retailers. The research suggests that as of 2020 only 45 per cent of businesses were even aware they can change their water provider. This can lead to potential issues, such as delays in identifying and contacting the relevant party in charge to resolve a particular issue or complaints submitted to CCW naming one party instead of the other as responsible for an issue.

Different parties may have different views about where the boundary between the retailer and wholesaler should be. Some retailers may feel wholesalers are encroaching on retail activities and risk undermining the business case for offering a retail service, whereas other retailers may be keen to avoid the costs of providing certain services that wholesalers could also provide. Most customers are likely to be more concerned that the service is provided in a timely, efficient and minimally disruptive way rather than worrying if it is the retailer or wholesaler who does the work. When a customer raises an issue about service provision, wholesalers and retailers should work collaboratively to respond to the issue raised rather than directing the query from one party to the other without addressing the issue identified.

2.3 Influences on wholesaler engagement with end customers

Unlike retailers, wholesalers do not normally face the risk of non-household customers switching to a rival wholesaler if they provide a poor or inadequate service.¹¹ The benefits associated with competition incentivising firms to improve their service offering or risk losing market share are not in place. But they do face some regulatory measures that may influence their conduct in the business retail market and/or affect the experience of non-household customers.

2.3.1 Market Performance Framework

The MPF comprises a set of processes and methods aimed at ensuring the orderly operation of the business retail market.¹² The MPF consists of two sets of financial metrics:

- the Market Performance Standards (MPS), setting out the different levels of underperformance for various market processes, and
- the Operational Performance Standards (OPS), which set out the level of performance wholesalers are expected to achieve regarding their operations and services.

Overall the MPS and OPS focus more on wholesaler-retailer interactions, and includes targets around the timeliness of processes. It is not directly concerned with the end customer experience, although meeting the targets around prompt service is likely to be valued by non-household customers, all else being equal. The MPF roadmap highlights the need to strengthen the links between the MPF and customer outcomes.¹³

In addition, Ofwat has also emphasised that while wholesalers generally comply with Market Codes (including the MPF), they need to be more proactive in terms of understanding the impact of the services provided to end customers, and tailor these to better reflect the needs and preferences of end customers. This further

¹¹ The largest customers, though, may decide to become self-suppliers or switch to using a NAV.

¹² MOSL: Code Subsidiary Document No. 0002: Market Performance Framework [\[online\]](#).

¹³ MOSL, Economic Insight and MAC Panel & Committee (2019): "Roadmap for the Evolution of the Market Performance Framework" [\[online\]](#).

supports the development of a qualitative measure of wholesaler performance to address the issues identified around wholesaler service provision.¹⁴

A recently introduced retailer measure of experience (R-MeX) provides reputational incentives to improve retailer-wholesaler interactions. A league table summarising the average score received by each wholesaler across seven areas of service is published as part of the measure.¹⁵ This measure addresses some of the concerns with more mechanical quantitative measures of service offering, and is intended to capture how well wholesalers are doing in providing the more intangible and qualitative aspects of good service. However, the measure is concerned with retailer experience and not directly with the non-household business customer.¹⁶ At best, there may be an indirect link with wholesalers more likely to receive a low R-MeX score if retailers have lots of customers unhappy because of things that were partly or wholly the responsibility of the wholesaler.

2.3.2 PR19 performance commitments and measures of customer experience

In terms of the price control framework, the PR19 Outcomes framework provides incentives to the 17 largest water and wastewater companies operating in England and Wales around the level of service provided through performance commitments and outcome delivery incentives (ODIs) that are both common to the industry and specific to companies. For example, common performance commitments cover areas such as supply interruptions, leakage, sewer flooding, and asset health metrics such as mains bursts. The performance commitments specify the level of service companies are expected to deliver and the ODI rates determine any payments (under- or outperformance) associated with the different levels of services delivered.

In terms of customer experience, PR19 introduced both a customer measure of experience (C-MeX) focusing on the experience of residential customers and a developer measure of experience (D-MeX) focusing on developer services.¹⁷

The idea of a B-MeX incentive for business retail customer experience was also considered, and Ofwat decided against introducing such a measure at the time. Ofwat did not consider it appropriate to set a direct measure of business customer satisfaction for companies operating in England as in Ofwat's view the business retail market and the market code framework should be the first avenue to explore for introducing this measure. Nonetheless, Ofwat did introduce bespoke performance commitments focusing on business customer experience for companies located wholly or mainly in Wales due to the differences between the English and Welsh retail markets.¹⁸

2.3.3 Potential gaps in current industry framework to incentivise wholesaler performance for business customers

From our review of the activities and services provided by wholesalers in the business retail market and current regulatory arrangement, we identify two key thematic problem areas a B-MeX incentive may address.

- First, the potential gaps in the current industry framework mainly relate to qualitative aspects of wholesaler-led end-customer interactions. These could cover: the experience of business customers having wholesaler staff on site to deal with the maintenance of assets / service failures; their experience arranging an appointment; the degree to which the wholesaler completes the job in an effective and efficient way; and the consequences of the visit (e.g. passing old and new meter reads to the retailer and

¹⁴ Ofwat (2020): "Review of incumbent company support for effective markets" [\[online\]](#).

¹⁵ The R-MeX league tables are published [online](#).

¹⁶ Ofwat (2020): "Wholesale Retail Code Change Proposal – Ref CPW084" [\[online\]](#).

¹⁷ Ofwat (2019) "PR19 draft determinations: Delivering outcomes for customers policy appendix" [\[online\]](#).

¹⁸ Ofwat (2019) "PR19 draft determinations: Delivering outcomes for customers policy appendix" [\[online\]](#).

customer where the wholesaler replaces a meter). How well the wholesaler delivers on such factors typically involve judgement and they are not easily quantifiable.

- Second, the current MPF relates to market and operational processes, while the R-MeX measure focuses on qualitative aspects of wholesaler-retailer interactions. The MPF lacks the links between the services provided by wholesalers and customer experience

Wholesalers also need to be clear about their responsibilities towards end customers and not inappropriately refer them back to the retailer.

While it may be difficult to quantify objectively how well a wholesaler has interacted with a non-household customer, in other market settings this would be something that firms subject to competitive pressures would be keen to do well for fear of losing customers to their rivals. Customers who consider a firm's staff to have been rude are more likely to switch to rival suppliers, even if they cannot exactly explain why they were unhappy with the service beyond stating that they considered the staff rude. Similarly, a customer who has been passed back and forth between the wholesaler and retailer because the wholesaler has not been clear about their responsibility / role towards end customers could also lead to the customer switching suppliers in a competitive setting. A well-designed B-MeX regime could potentially provide wholesalers with similar incentives to improve the customer experience.

3 Is It Practical to Design a B-MeX Incentive?

This chapter considers at a high-level the practicalities involved in designing a B-MeX incentive that might address the gaps in the current regulatory framework identified in Chapter 2. We review the lessons that might be learned from regulatory precedents both in the water sector and elsewhere. We also consider some of the questions around the scope of activities that may be covered in a B-MeX regime, how a B-MeX score might be generated, other key policy issues to consider and the criteria to use when assessing options.

3.1 Lessons learnt from the sector and other sectors

As a first step to thinking about how a B-MeX regime might be designed, we have considered approaches that regulators have used in similar situations. The exercise illustrates that the seemingly simple exercise nevertheless requires a number of policy decisions. The appropriate choices will depend on the relevant policy objectives and the specific features of the sector.

Our review of the potential lessons learnt in terms of designing an incentive mechanism focusing on business customers' satisfaction considered a range of sectors. This included precedents from:

- the water sector (the Scottish business retail market);
- regulated sectors with a business retail market (electricity); and
- sectors where the services delivered to consumers and therefore the end-customer experience depends on two (or more) suppliers (rail and aviation).

Table 2 below summarises some of the key lessons from these sectors in terms of the presence of a quality of service regime (typically included within the price control framework), whether there exists a separation between residential and business customers, whether satisfaction surveys are used to measure customer experience in terms of the services provided, and whether direct compensation is available to end customers (including outside a price control setting). The Appendix provides further details on the quality of service regime used in each of these sectors.

Table 2: Summary of lessons learnt from other sectors

	Scottish business retail market	Aviation	Electricity	Rail
Customer experience depends on wholesaler-retailer type interaction	✓	✓	✓	✓
Quality of service regime (within price control framework)	✓	✓	✓	✓
Separation between residential and business customers	✓	✗	✓	✗
Customer satisfaction survey	✓	✓	✓	✗
Direct compensation to customers (incl. outside price control framework)	✗	✗	✓	✓

While there are lessons to learn from other sectors, there is no single comparator that exactly mirrors the situation in the water business retail market. While many of the high-level features of the Scottish business retail market are similar to the English market, important differences include the presence of only one wholesaler in the Scottish market. This means that a relative performance scheme is not possible, whereas it may be possible in England.

3.2 Basic design of a B-MeX incentive

3.2.1 Scope and scale of wholesaler activities covered by B-MeX incentive

There are various approaches to determining the scope of the activities to be covered by a B-MeX incentive.

- The B-MeX incentive to **cover all the activities** which are related to the direct interaction between wholesalers and business customers. In this case wholesalers would be incentivised to improve on all related aspects impacting the end customer experience. In cases where the scheme covers a broad range of activities, consideration will need to be given to the weights given to different activities included. If equal weights are given to all activities, wholesalers may decide to improve the activities that are easier and less costly to improve while focusing less (or not at all) on issues which could potentially be of higher importance to customers.
- The B-MeX scheme could **focus only on a few specific activities**. In this case wholesalers would have less scope and incentive to focus on improving the least costly activities at the expense of activities that are perhaps more important to customers. This approach may be less suited if there were any significant differences between the services provided by wholesalers across different regions. In addition, ensuring that the incentive covers the experience of customers in regions with fewer business customers will equally be important. For example, in areas with fewer business customers, the likelihood of customers facing the subset of issues covered by the B-MeX incentive focusing on a few specific activities is likely reduced. Consequently, this could mean that less information will be captured by the B-MeX scheme in terms of end customer experience, which could risk companies facing the financial payments attached to the incentive based on a small number of interactions.
- There could also be a **different B-MeX incentive (or score) for each activity** within the scope of the incentive. An advantage of this approach would be that companies will be incentivised to improve performance for each of these activities. One clear disadvantage of this option relates to the set-up and monitoring costs of the B-MeX measure(s) which are likely to be disproportionately large compared to the improvements in end customers outcomes. This is because for each B-MeX incentive (or score) the design issues explored in later chapters (e.g. applying different weights, using relative or absolute measures of performance, etc.) would equally need to be considered.

The final choice of activities to include will entail decisions about how to treat activities where end-customer experience partly depends on the behaviour of retailers. A basic challenge in the design of B-MeX incentive is ensuring that it creates incentives for wholesalers to provide a better service for end-customers, with any rewards and penalties accruing to wholesalers being linked to their performance.

At first glance, it may seem unfair to penalise or reward a wholesaler for something that is not entirely their responsibility. However, we would caution against adopting a B-MeX regime that only has financial or reputational incentives for services where no responsibility could be attached to the retailer. A B-MeX regime is about incentivising the wholesaler to do their part to realise better outcomes for customers. In some cases, it may be possible for the wholesaler to work with the retailer to improve how both parties engage with business customers.

Moreover, B-MeX scores will inevitably have some 'noise' and in any given period some wholesalers will receive scores higher or lower than their efforts to provide a good service warrant. For example, the B-MeX

measure may use a survey and the sample in a particular month could select a relatively large number of business customers with a propensity to give low scores. In this and similar cases, the B-MeX regime should not and could not correct for all these idiosyncrasies. It is working as intended if wholesalers are incentivised to make an appropriate effort to improve the service they offer to business customers and in the long run the vagaries of chance will even out. If there are some activities where their behaviour is partly, but not fully, responsible for the customers' level of satisfaction, B-MeX should be incentivising the wholesaler to attempt to improve their service offering, not giving a message that there is no need to do anything because retailers also have a role to play.

Another important issue is how to deal with activities that are provided to end customers by some wholesalers but not by others when designing a B-MeX incentive. An obvious example would be services relating to sewage that water-only companies will not have to address. In other cases, the null return from a region may be temporal (e.g. no engagements relating to sewage). How should a null return be treated in the B-MeX incentive regime? Should full marks be given, or should this aspect be excluded from the calculation of a B-MeX score (and any financial payments attached)? There would be perverse incentives if a wholesaler was deterred from proactively remedying a problem because it would no longer qualify for a B-MeX reward. Against that, the specific behaviour B-MeX is focused on is how the wholesaler deals with business customers, and the incentives are intended to reward or penalise firms for how they do this. Moreover, in many cases the absence of activities in a given region will be due to exogenous factors rather than any specific actions that the wholesaler took to pre-empt the need to provide a given service in the business retail market.

Currently the available evidence regarding what business customers value the most in terms of the qualitative aspects of the activities and services provided by wholesalers is limited and almost surely not an adequate information base to finalise a B-MeX scheme. Data on complaints collected by CCW could provide some initial guidance around the scope of activities to be covered by the B-MeX incentive, but we would not recommend relying solely on this information for the scheme. Besides CCW data, some individual wholesalers have commissioned their own research looking at the activities and services customers value the most. For example, a wholesaler has commissioned research to better understand the experience of non-household customers with the company, any improvements suggested in terms of the activities and services provided, as well as to assess customers' understanding of the business retail market. However, these surveys investigating non-household customer preferences and valuations were not specifically undertaken with a view to designing a B-MeX regime and they may not be appropriate to use if trying to determine B-MeX incentives for other wholesalers given the possibility of regional differences in customer preferences.

3.2.2 Generating a B-MeX score

In addition to identifying the activities to be covered by a B-MeX regime, it is also necessary to think about how to measure wholesaler performance in these activities. Potential options include:

1. a quantitative measure (score) based on certain activities carried out by wholesalers in a timely way;
2. using complaints data relating to the services provided to end customers;
3. developing a customer survey focusing on qualitative aspects of service provision; or
4. some combination of the first three options.

Given the aim for the B-MeX incentive to focus on improving end customer experience, relying exclusively on quantitative measures is unlikely to adequately capture what is important to customers and will consequently create poor incentives. Customer feedback is a crucial component. We discuss the possibility of using complaints data in Chapter 5.

In terms of the use of a survey, a key decision is who gets to provide feedback. We envisage two broad possibilities for the type of survey that could inform B-MeX. The first would be to **survey a sample of all**

business customers, samples to ascertain their satisfaction with the wholesaler. We would not recommend this option. It is likely to result in a large percentage of the businesses surveyed being relatively uninformed. Many may not be aware of the respective roles of the retailer and wholesaler. Moreover, most are unlikely to have directly interacted with a wholesaler in the past year.

Instead, we suggest that it will be better to limit any survey to customers that have had contact with the wholesaler. This could include

- Business customers who directly interacted with wholesaler (e.g. meter repair, flooding), asking the customer how satisfied they were with the service provided.
- Business customers who communicated with wholesalers (e.g. around leakage allowance, other requests) asking customers how well the policy was communicated to them, or how their request was handled.

Both of the above options under the second approach above would capture only those customers that have had contact with the wholesaler, which is similar to the C-MeX customer service survey. Given that not all business customers would come into contact with a wholesaler in a given period, these survey options may in practice survey a subset of business customers whose characteristics may differ from the non-household customer population more widely. There will be an implicit weighting towards customers who interact with wholesalers. We explore whether there would be merit in making further adjustments to the weights attached to the responses of different customers in the next chapter.

3.2.3 Other key policy issues for B-MeX incentive

In addition to the issues around the type of survey and activities to be covered by the incentive, there are also a large number of important decisions that need to be addressed when designing the B-MeX scheme. In subsequent chapters of this report, we discuss some of these key design decisions that need to be resolved:

- Weighting of customers;
- Use of complaints;
- Absolute or relative performance measures;
- Regulatory instrument to use; and
- Size of financial incentive.

In undertaking this assessment, we have had regard to three important criteria, agreed with Ofwat, CCW and MOSL as part of this study.

- Consumer interest – *Does it make (individual) customers better off?*
- Effectiveness – *Does it create the right incentives for wholesalers to improve business customer experience given the wholesaler-retailer interaction (while minimising the risk of unintended consequences/ regulatory failure)?*
- Cost/ practicality – *Is it proportionate, practical and easy to enforce?*

Where it is potentially material, we have also considered factors such as how a given policy choice might affect other regulatory incentives (e.g. would the B-MeX regime undermine other incentives that Ofwat wants to create), whether it would account for any material regional differences, and which services could be covered by the B-MeX regime.

4 Weighting of Customers

4.1 Introduction

The design of a B-MeX incentive would need to appropriately consider whether the experience of certain customers should have more weight than those of others.

This chapter discusses the issues associated with weighting customer responses in the B-MeX survey. A background to the business customers in the water sector is provided in subsection 4.2.1. We then describe the weighting issues in market research more generally in 4.2.2. Issues of weighting business customers by size are then discussed in 4.2.3, and additional considerations as to how size is defined in 4.2.4. Issues of weighting customers on other measures are raised in 4.2.5.

Subsection 4.3 offers recommendations of whether to adjust customer weightings in the B-MeX survey.

4.2 Description and analysis of issues

Weighting customers could present a range of issues in terms of their effectiveness in incentivising the wholesaler to improve service quality, the practicality of implementing the weighting, and the ability to capture different customers' views.

There are two broad issues that arise in considering the decision to apply different weightings to different customers in B-MeX:

- Weighting within market research to ensure that reported survey results are representative of the population. This is a measurement issue.
- Weighting within B-MeX, which is over and above the first issue, and which is about which customers we want wholesalers to focus on most. This is a policy issue.

We discuss each of these issues below.

4.2.1 Background: business customers in the water market

Whilst there is considerable diversity in business customers, a minority of large customers are responsible for the majority of the water usage of business customers. In 2019-20 large businesses with more than 250 employees account for approximately 64 per cent of water consumption and their average annual expenditure is 100 times that of the smallest business category, microbusinesses, with less than ten employees (see Table 1).

In the “State of the market report” published by Ofwat, business customer size is based on the **number of employees** for the purpose of presenting the distribution of business water customer characteristics. This is one plausible approach to defining business customer size among numerous other approaches. Indeed, the table above also shows that weighting by the number of employees would not reflect the actual **volumes of water consumed** by each customer – small and medium enterprises (SMEs) account for a full quarter of consumption but only 13 per cent of customers. Similarly, size could be defined on the basis of **water expenditure**. Alternatively, it may be beneficial to capture the **number of premises** of different business customers in the definition of size. With equal customer weighting, the feedback from customers with multiple premises (and who may consequently be more sensitive to poor wholesaler service) could count for as much as those with just one.

We revisit the advantages and disadvantages of each of the aforementioned approaches after first discussing some more general issues with weighting by size.

4.2.2 Issues with response weightings in surveys

This subsection briefly discusses the first issue related to the decision to apply different weightings to different customers in B-MeX, and demonstrates the relevance of response weighting to ensure representation in a B-MeX survey in which implicit customer weightings might occur without it.

Weighting within market research to ensure population representation

If the B-MeX incentive is based on survey responses, any decision to weight businesses by size would need to consider the expected response rates of businesses of different sizes. In Ofwat's non-household customer survey (see box below) smaller customers have lower response rates relative to their number in the market. Weighting within market research is typically employed to ensure that reported survey results are representative of the population, rather than of the responding sample. Examples of this in the regulated sectors are provided in the box below.

Box 1: Weighting customers to improve customer representation in customer market research

Weighting business customers to make the pool of respondents more representative of the population is common in survey design and has been applied in, for example, Ofwat's non-household customer insight surveys and Ofgem's surveys of micro and small business engagement surveys (which are not for incentive purposes).

Ofwat's 2020 non-household customer insight survey significantly reduced the weight applied to large customers (defined as those with more than 250 employees) and significantly increased the weight applied to microbusinesses (with fewer than 10).¹⁹ This was to ensure that overall results for the market reflect the distribution of customer size in the market.

Ofgem's 2016 survey of micro and small business engagement involved adjusting the survey response data to reflect the actual proportion of micro and small businesses with non-domestic energy contracts.²⁰ It used Random Iterative Method ("Rim") weighting which adjusts multiple characteristics of the sample – in this case, size and sector – simultaneously to provide a distribution of characteristics proportionate to the population. These proportions were established in 2014 and have been held constant.

The approach taken in the Ofwat and Ofgem customer surveys (and other surveys) addresses a measurement issue in survey design. Any B-MeX survey design should consider similar weighting approaches given the potential for different response rates across customer types and the resulting implicit weighting this could cause.

Implicit weighting in the B-MeX survey

In the preceding chapter, we recommended that a survey of non-household customers should be confined to those business customers who have directly interacted or communicated with wholesalers. This implicitly has implications for the weights attached to different customers' experience.

For instance, taking the list of all wholesaler interactions and surveying a random sample would implicitly weight by the number of wholesaler interactions. From the discussion above, it is likely that such an implicit weighting would favour larger customers (who tend to interact more with wholesalers relative to their

¹⁹ BMG (2020) "Non-household Customer Insight Survey 2020 BMG report", a report prepared for Ofwat and CCW [\[online\]](#).

²⁰ Quadrangle (2017) "Micro and small business customer engagement in the energy market, 2016", report appendices, a report prepared for Ofgem [\[online\]](#).

number in the market), and that this could have the benefit of incentivising wholesalers to focus on these larger customers (thus mimicking some of the incentives in a competitive market).

Alternatively, B-MeX could take a list of customers that have experienced a wholesaler interaction, such that each customer appears once on the list, even if they had multiple interactions. Taking a random sample from this list would also implicitly weight by wholesaler interactions because it affects the likelihood that different types of customer will end up on the list.

Therefore, without actively changing the relative weightings applied to different customers, the above sampling options would imply different customer weightings. Adjustments to correct for the possibility that certain types of business may be less likely to respond to a survey when selected could ensure that the B-MeX score would correct for the possibility that the sample of wholesaler interaction was unrepresentative of the population of interest. But whether further adjustments are warranted would be a policy decision about which customers wholesalers should be incentivised to focus on the most. If the starting population contains, for example, relatively more large businesses (as it is likely to if it includes all wholesaler interactions), then even if all of the selected sample responded to the survey, there may be policy reasons for deciding to make weighting adjustments.

4.2.3 Issues of weighting business customers by size beyond the statistical representation weights

There are advantages and disadvantages of weighting larger business customers more highly.

Assigning higher weights to larger companies may be beneficial for the purposes of eliciting feedback from business customers which relates exclusively to the service provision of wholesalers (and not retailers). Larger users may be more capable of distinguishing the wholesaler's service from that of the retailer, and therefore might be able to provide a more informed perspective on the quality of their wholesalers' service provision. This may be because larger firms have internal teams dedicated to their utility supply. As of April 2020, 96 per cent of large customers were aware they can choose their retailer, compared with 58 per cent of all business customers.²¹ This difference in awareness could signal that larger customers with better knowledge of the market might be more capable of distinguishing the retailer from the wholesaler, and thus able to more accurately report on their experience with the wholesaler. However, motivating higher weights on larger companies on the basis of market awareness would penalise smaller firms for their lack of it, since they are unlikely to command the same resources for maintaining current knowledge on the market. Given this, it may be more appropriate to weight smaller customers more highly precisely because they are less well informed about the market.

Different problems might be experienced by users of different sizes. Weighting by customer size may implicitly be a weighting based on types of issues. If larger customers were weighted higher, then experience with issues predominantly experienced by them would count for more than the issues predominantly experienced by smaller customers. Yet large businesses comprise less than 1 per cent of all business customers, whilst microbusinesses comprise 86 per cent.

Another issue is whether customer weighting based on size would require adopting bespoke B-MeX schemes for individual water and wastewater companies, which may run contrary to other policy goals. This issue was flagged in Ofwat's discussion of whether to weight customers by size in the developer customer satisfaction survey of the D-MeX in the PR19 draft determinations.²² There, water companies were asked to identify the size of each of its customers, defined in terms of the number of connections. The observed significant variation across companies in the size of their developer services customers meant that weighting customers

²¹ Ofwat (2020) "State of the market 2019-20: review of the third year of the business retail water market" [[online](#)].

²² Ofwat (2019) "PR19 draft determinations: Delivering outcomes for customers policy appendix" [[online](#)].

based on size would in practice cause different weightings for each *water company*. This was considered to be inconsistent with Ofwat's non-water company specific approach to D-MeX,²³ whereby the same approach applies to all water-only companies and water and sewage companies. For B-MeX, it would be important to know whether there is an observable correlation between the wholesalers and the size of their respective business customers.

Wholesalers would have an incentive to focus their attention on customers with the highest weights (i.e. the larger customers). This would be the case if the wholesaler has accurate information on the customer. For example, if a wholesaler knows that a particular group of customers are significant water consumers, it might prioritise the resolution of these customers' problems. Provided that the weights are appropriate, then arguably this approach would align with the incentives present in a competitive market on firms to focus service quality on customers from whom they receive the highest income.

Issues also arise with the treatment of *self-suppliers*. Although there are currently only 12 such customers (as of March 2020), self-supply may be an attractive option for larger business water customers.²⁴ Self-supply customers deal directly with the wholesaler and are perhaps more likely to make contact with it in any given period. They may also have a richer understanding of the wholesaler's service quality – perhaps more so than other large customers. Consideration could thus be given to weighting self-supply customers more highly than larger customers. On the other hand, self-supply customers represent an extreme case for business customers, and their water requirements are unrepresentative of the needs of other business customers. The issues on which they report may therefore be unrepresentative of the business customer market and it would be undesirable to design a B-MeX regime that created incentives for wholesalers to skew their attention to this particularly small group of customers.

Finally, weighting customers by size could send a signal that larger customers are more important to society if size were understood to be a proxy for importance. Aside from being relatively arbitrary, this could put start-up firms and other smaller customers at a relative disadvantage if it meant that their water and sewerage supply issues are not addressed as well as those experienced by the larger incumbents.

The merits and drawbacks of weighting responses from larger customers more highly are summarised in Table 3 below. The table shows the expected performance of weighting by size against three agreed criteria (effectiveness, cost/practicality and customer interest) as well as any other considerations, against the counterfactual of not applying weights beyond the statistical representation weights. It is a somewhat circular task to assess the performance against the third criterion, customer interest, given that the guiding principle in increasing the weights on certain customers is to determine the relative importance of the interests of different customer groups. Therefore, the way in which different customers are weighted consequently affects the outcome for these customers.

²³ The D-MeX policy decision document [\[online\]](#) noted no “evidence from the pilot of the need to apply company-specific or other adjustments to D-MeX, as the impact of these variations on the services covered by D-MeX appears to be minor. We are therefore not minded to treat companies differently” (p.24).

²⁴ Ofwat (2020) “State of the market 2019-20: review of the third year of the business retail water market” [\[online\]](#).

Table 3: Summary of performance of weighting by size (positive points denoted by “+”, negative by “-”)

Criteria	Increasing the weight given to responses of larger customers
Customer interest	<ul style="list-style-type: none"> + Representative of those with highest demand for water. + Customers (including smaller customers in terms of employee count) are more likely to receive service quality that matches what they pay. + Representative of customers that could be exposed to more issues, including self-suppliers. - The incentive to put more effort into resolving issues of larger customers could be to the detriment of the vast majority of the market of smaller firms.
Effectiveness	<ul style="list-style-type: none"> + Captures customers with more awareness of water market and can accurately report on their experience with the <u>wholesaler</u>. + Wholesaler has incentive to be receptive to service quality issues of customers that spend the most, which aligns with incentives in a competitive market. - Size weighting may, in practice, amount to weighting by type of issue. - Penalises smaller firms.
Cost/practicality	<ul style="list-style-type: none"> + Data on customer size metrics (consumption, expenditure, employee count etc.) are generally available to the regulator and may not require additional collection. - Identifying sufficient responses for a particular size category in each wholesaler region could be difficult.

4.2.4 Further issues related to the way in which “size” is defined

If weightings were to be applied to business customers, other issues arise from the way in which “size” is defined.

An initial way of categorising customers could be by the **number of employees**. These data are currently recorded for the business retail water market and inform the size categorisation in the “State of the market” reports. This method would assume that customers within the same size category have the same level and type of water service requirements. Thus, with this method the feedback from customers with significant water usage would count for the same as those similar sized customers with more limited water or waste water requirements. The method would not distinguish between customers based on whether they use a given wholesaler for water, wastewater, or both. Weighting customers by the number of employees is therefore considered to be an inappropriate method of characterising large water customers.

Other methods would capture the actual quantity of water consumed by different customers, such as **weighting by the volume of, or expenditure on, water consumed**. These approaches would capture the diversity in expenditure of customers within each size category (based on employees); the annual bill for water services of SME customers varies from £500 to more than £100,000 (with an average spend of £2,500), and the range is similar for the largest customers (with an average spend of £35,000).²⁵ If size weights were not based on expenditure, the experiences of some smaller customers with especially high expenditure (and which may feel entitled to a louder voice) would count for the same as those that pay a fraction of that amount.

Weighting by the number of premises of each customer could be an option if the B-MeX survey is not designed to allow a customer to report on the experience of each of its premises individually. It could better capture the experiences of customers which have multiple connections to the water network and have potentially higher exposure to the service provision of the wholesaler. It could also help to represent self-supply customers with multiple premises. It is less sensitive to mergers: with equal customer weighting a merged entity whose premises were combined would account for less than the sum of the pre-merger weights of the separate organisations, whereas weights based on premise numbers would be proportionate

²⁵ Ofwat (2020) “State of the market 2019-20: review of the third year of the business retail water market” [[online](#)].

to the number of premises controlled by the merged entity. A downside of this approach is that customers with the same number of premises might have different usage / water requirements (e.g. a hotel chain is likely to have very different requirements from a retail store chain).

4.2.5 Other approaches to weighting customers

Given the issues related to weights based on customer size, it is worth considering other weighting approaches. A non-exhaustive list is as follows:

- **Weighting by the number of interactions or contacts with the wholesaler.** This would involve customers that have had the most interactions with the wholesaler being weighted more highly. This approach would best represent customers with particularly frequent need for wholesale services. It would thus reflect the actual contacts with the wholesaler that customers have experienced, regardless of their size. The regulator would need to define “interactions” or “contacts” to avoid inflation of weights based on superficial interactions. The approach would arguably account for the full range of issues that the wholesaler is responsible for (those of small and large customers, of water and wastewater customers, and of customer with multiple premises) to the extent that they are reported on by the customer. However, larger customers might have sufficient resources to provide feedback on each interaction, whilst smaller customers might not.
- **Weighting by issue or type of interaction with the wholesaler.** This would require the regulator to decide on the relative importance of different types of issues experienced by business customers – an exercise which could require regular recalibration and could therefore be costly. This brings the risk that the issues weighted highly are not those where customers most value. The ranking of wholesaler issues could potentially be informed by complaints, which may correlate with the impact that customers care about most. It could, in turn, be useful for calibrating the B-MeX incentive to encourage wholesalers to direct resources towards addressing certain issues by increasing the weights on them. Ofgem considered attaching different weights to responses to a customer satisfaction survey depending on the “milestone” of the customer journey with electricity transmission connections. It ultimately decided that each customer response from each survey milestone should hold the same weighting and be actioned by electricity transmission operators on its own merit (see box below). A related issue is whether it would be more appropriate to have separate B-MeX customer satisfaction scores for different types of issues.

Box 2: Weights on milestones in consumer journey with electricity transmission operators in RIIO-2

For the upcoming control period, RIIO-2, Ofgem introduced a financial incentive based on a survey aimed at measuring the quality of the electricity transmission connections process.²⁶ This involved establishing “milestones” in the connections process (e.g. pre-application engagement, project delivery) where each milestone triggers a customer survey.

Ofgem decided upon weighting responses received at each milestone equally.²⁷ This was considered to ensure that the electricity transmission operators would action each survey response on its own merit. However, some draft determination consultation responses raised the issue that equal weighting could “skew the overall survey score in favour of large corporate entities”.²⁸

- **Weights based on whether customer receives water, wastewater or both.** Insight from the C-MeX incentive suggests that water companies that provide wastewater services receive higher customer

²⁶ The Quality of Connections Survey (GCS). See pp. 23-29 of Ofgem (2020) Final Determinations ET Annex REVISED [\[online\]](#).

²⁷ Ofgem (2020) “RIIO-2 Draft Determinations - Electricity Transmission Annex”, Appendix 2 - Quality of Connections Survey Methodology [\[online\]](#); Ofgem (2020) Final Determinations ET Annex REVISED [\[online\]](#).

²⁸ Ofgem (2020) Final Determinations ET Annex REVISED, p.28 [\[online\]](#).

satisfaction scores on average (see box below). To ensure that smaller, water-only companies are not penalised in any relative comparisons, it may be appropriate to try and weight on this basis. It would account for the possibility that survey responses differ systematically based on the type of water service. This approach would not be necessary if there were two separate measures for water and wastewater.

Box 3: Weights on water and wastewater satisfaction scores in C-MeX

At PR19, Ofwat put equal weights on water and wastewater satisfaction scores within the operations component of the C-MeX customer service survey (25 per cent each).²⁹ This was considered unfair by water-only companies because, first, wastewater satisfaction scores are typically higher than water scores and, secondly, there are around twice as many water-related contacts from customers as wastewater. The latter point would, with equal weighting, up-weight the wastewater contacts in practice.

Overall, equal weighting was expected to cause higher absolute scores for water and sewerage companies, and higher scores relative to water-only companies. However, because the customer service satisfaction score contributes only 40% per cent of the C-MeX score, the impact of choosing to apply different weights to water and wastewater contacts caused only minimal differences to the overall C-MeX score.

- **Weighting by level of activity.** Ofwat collects information on the level of activity of business customers, where activity refers to business customers switching retailer, re-negotiating a deal with their existing retailer, or actively considering or trying to switch or re-negotiate. A rationale for this approach would be to capture the experience of business customers that are more aware of the market and the relationship between themselves, the retailer and the wholesaler, and in turn provide accurate reflections on their experience with the wholesaler. In practice, this method could amount to a weighting in terms of customer size because activity appears to increase in customer size based on the number of employees (see Table 4 below). Further, it would underrepresent customers that have been 'inactive' for reasons not linked to their use of wholesaler services. It would also give more weight to customers of those retailers where customer churn is highest.

Table 4: Proportion of business customers that were active in last 12 months (2020)

Customer size (number of employees)	Proportion active in last 12 months
Micro	8%
SME	8%
Small	6%
Medium	16%
Large	26%
All	8%

"Active" denotes all customers that have switched or re-negotiated, are actively considering it, or have tried to do so.

Source: BMG Research (2020) "Non-household Customer Insight Survey 2020 BMG report" [[online](#)].

- **Weighting by business type.** This approach would signal that the use of water by particular business customers counts for more than that of others. For example, more weight could be given to customers involved in the provision of public goods, such as hospitals, councils and leisure centres, or to different customers based on their industrial classification code. It is a very interventionist approach that would entail Ofwat making value judgements about what types of business are most important and therefore worthy of good wholesaler service.

²⁹ Ofwat (2019) "PR19 draft determinations: Delivering outcomes for customers policy appendix" [[online](#)].

4.3 Recommendations

This chapter has discussed the basis for assigning weights to business customers in the B-MeX survey.

We recommend that any B-MeX survey should consider including statistical adjustments to ensure that the results come from a sample population that is representative of the population being sampled. So if small companies are less likely to respond to a survey, the responses from those that do respond might be given more weight or a higher percentage of small companies may be surveyed. However, the population being sampled may not be the 1.2 million non-household customers, but instead the number of business customers that have had a recent wholesaler interaction, so it need not be the case that the statistical corrections mean that the weightings given to different business types correspond to their share of the business population.

We recommend not actively changing the weights assigned to business customer responses in the B-MeX survey beyond the statistical representation weightings that are typically applied in market research. There are reasons that could justify such intervention. For example, equity considerations may prompt special concern for smaller customers. Alternatively, there may be a belief that giving more weight to large customers will better mimic competitive markets by giving water companies incentives to focus service provision on key customers who consume the most or a desire to give more weight to 'informed customers' more able to distinguish between the services of the wholesaler from the service of the retail.

However, on balance we do not think such an intervention would be warranted. A B-MeX survey based on customer interactions will implicitly be weighted to larger non-household customers. Further increasing the weight given to such customers would generate a number of disadvantages:

- It may implicitly amount to a weighting based on types of issues, where wholesalers may have less incentive to resolve the issues typically faced by the vast majority of (smaller) business customers.
- It risks tying importance to any particular issues faced by larger companies, the resolution of which by the wholesalers may not benefit business water customers in general.

Decreasing the weight given to large customers is also of questionable value. It would not be consistent with giving water companies the same incentives to provide good service every time they interact with a non-household customer. It would mean that responses from potentially better-informed customers (on the basis that they are more engaged in the market and have more experience to draw on) would be given less weight. This leaves aside the practical issues associated with attempting to determine the appropriate weights.

5 Use of Complaints

5.1 Introduction

Complaints were used as part of the previous service incentive mechanism (SIM) and are currently used as part of the current C-MeX incentive.

In this chapter we consider two broad ways that complaints data might be used for B-MeX.

1. They could help in identifying the issues that are of concern to customers, and their relative importance.
2. They could be used as a component of the B-MeX scores, such that wholesalers have reputational and possibly financial incentives to minimise the number of complaints.

One other possibility would be to use complaints data to identify customers to include in surveys used to generate B-MeX scores. We do not think this possibility has much merit. There are simpler ways to develop a sample for surveys. Moreover, there are obvious selection bias concerns, as the sample would be skewed to dissatisfied customers.

5.2 Description and analysis of issues

When thinking about whether and how to use complaints data, important considerations will be: the data that are currently available, the reasons why people might make complaints, and the incentives for changed behaviour that might arise if complaints data were used in the design of a B-MeX regime. We discuss all of these ideas in the sections that follow.

5.2.1 Currently available complaints data

Complaints data that are already collected have the obvious practical advantage that they could be used in a B-MeX scheme without the expense of designing surveys or arranging for additional data collection.

CCW receives non-household complaints directly and also records customer complaints received directly by retailers (directly by wholesalers in Wales). Complaints directly to CCW currently represent a minority of the total market complaints. In its most recent report documenting complaints, CCW reported 14,363 non-household complaints to retailers (and Welsh companies) and 3,436 complaints to CCW.³⁰

This represented a fall on the number of complaints received the previous year for both sources of complaints, the first time this had happened since the retail market was opened for competition in England in April 2017. Nevertheless, the volume of complaints remained higher than those recorded in 2016-17.

Since the opening of the retail market, wholesalers no longer interact day-to-day with non-household customers. The contractual relationship is between retailers and non-household customers. Non-households are expected to report complaints to their retailer and, if unhappy with the response, complain to CCW. There is no direct mechanism for complaining to the wholesaler. In 2017/18, CCW judged that seven per cent of non-household complaints against retailers that it received arose because of customer dissatisfaction with the wholesaler service.³¹

More recently, CCW has asked retailers to identify how many of the complaints they receive have some wholesaler element to them. This might relate to service failure by the wholesaler or a wholesaler policy that

³⁰ CCWater: “Non-household water customer complaints | April 2019 – 31 March 2020”, p. 26-29 [[online](#)].

³¹ CCWater: “Non-household water customer complaints | April 2019 – 31 March 2020”, p. 23 [[online](#)].

caused customer dissatisfaction. Retailers have reported that most written complaints that they receive are wholly attributable to the retailer, with less than a quarter attributable partly or wholly to the wholesaler. For example, in 2019/20 of 13,411 complaints by non-households to retailers, 942 were wholly attributable to the wholesaler and a further 1971 were partly attributable to the wholesaler and partly to the retailer.³² In total, retailers categorised complaints partly or wholly attributable to the wholesaler as accounting for 22 per cent of the total level of written complaints that retailers receive. These numbers align with what CCW would expect given its analysis of complaints to CCW. Similar data in 2018/19 was also consistent with the number of complaints wholesalers self-reported having been referred to them by retailers.

Our understanding is that the nature of the complaints relating to wholesalers has not changed significantly in the years since market opening. The concerns relate to operational issues, including leakage allowances.

Another potential source of 'complaints' data could come from the bilaterals transactions programme. MOSL is working with companies in the business retail market to improve the speed and quality of service that water companies provide to their non-household customers, including via the creation of centralised 'bi-laterals hub' through which requests for bilateral transactions are submitted and processed. The programme is intended to introduce a consistent approach for retailers and wholesalers communicating with one another. This should increase automation, creating more central data. It is possible that the bilaterals hub for example could be used to understand which bilateral transactions customers are undertaking and their satisfaction or dissatisfaction with the process, for example by requesting the end customer for feedback following the completion of the end customer's request for the bi-lateral transaction. It might also be possible to identify which wholesaler activities are most frequently impacting customers, at least to the extent that they prompt a retailer to make a bilateral transaction. However, an obvious problem with relying on these data is that it relates to communications from retailers rather than customers directly. It is possible that different retailers have different thresholds for raising a bilateral transaction. Given their differing market shares across regions and potentially changing market shares over time, that might create issues with making like-for-like comparisons across regions or over time.

5.2.2 Why people complain

Understanding why people make complaints is important to determining whether and how such information might be used in the design of B-MeX. Underlying the simple idea that people complain when the expected benefits of doing so exceed the expected costs are a myriad of different factors that might influence whether a non-household business actually complains. How representative complaints are of what concerns the generality of non-household businesses will depend in part on which of these factors are most important.

Europe Economics has previously identified a number of factors in the literature on consumer complaints that affect a consumer's decision to complain.³³

- **Degree of dissatisfaction.** Complaints are more likely to focus on instances where a customer is especially dissatisfied. Using complaints data to inform the design of a B-MeX scheme may lead to a focus on wholesaler activity that gives rise to particular consumer detriment. This may be considered desirable, as these are outcomes that non-household customers particularly want to avoid. However, it is possible that the complaints are outliers, and that there is little or no correlation with how different wholesalers interact with non-household customers generally. It is also possible that many instances where dissatisfaction levels are high enough to motivate someone to complain will involve factors above and beyond the services that B-MeX is supposed to capture. For example, a complaint about wholesaler

³² CCWater: "Non-household water customer complaints | April 2019 – 31 March 2020", p. 23 [\[online\]](#).

³³ Europe Economics (2007): "An analysis of the issue of consumer detriment and the most appropriate methodologies to estimate it" [\[online\]](#).

service being poor may be more likely if the wholesaler was responding to a flooded business premise or supply disruption than if they were repairing a meter.

- **Importance of purchase.** Customers are more willing to complain when the good or service is more important to them. Complaints may be more likely to come from large users of water and businesses where water supply is fundamental to their operations. A B-MeX scheme based on complaints data may, implicitly, be giving more weight to non-household customers for whom the wholesaler service is relatively more important. The desirability of this is discussed more fully in chapter 4.
- **Opportunity to complain and knowledge of process.** Complaints are more likely when it is easier for the non-household customer to make a complaint. A B-MeX scheme that used complaints data would need to satisfy itself that there were not significant differences across wholesaler regions in the ease with which businesses can make a complaint. Given the market shares of retailers are very different across different wholesalers, care would be needed to ensure that B-MeX scores based on complaints were not just identifying those wholesalers where the largest retailer is particularly good at informing customers about their right to complain.
- **Probability of complaint success.** Complaints are more likely where there is a belief that the complaint will be successful and either resolve the problem or lead to satisfactory compensation (monetary or psychological). This may mean that written complaints are more likely in instances where the complainant believes that the issue has not yet been resolved, as opposed to instances where the wholesaler provided poor service but completed the work. A B-MeX scheme which provided direct compensation to the affected business may also prompt more complaints than a scheme where any financial penalties did not directly compensate the complainant. Past experience with complaints and how they were handled may also have implications for the willingness of business customers to complain in the future. If past complaints led to a good outcome in the opinion of the non-household customer, they may be more likely to complain in the future. This might mean that complaints data come to be dominated by a few repeat complainers.
- **Personal characteristics.** Factors such as age, gender, income and level of education may all influence an individual's propensity to complain. Past studies have found that men, people aged 40-54 years old, the better educated, higher income earners, and more politically committed and liberal are more likely to complain. This may have implications for the mix of non-household businesses that actually complain, with more complaints from sectors with a relatively larger share of the workforce having personal characteristics that correlate with propensity to complain.
- **Situational influences.** Factors such as a liquidating the company, relocating the office, or a peak in demand for the company's services may all discourage businesses customers from complaining. Similarly, other people can influence the propensity to complain, either encouraging or discouraging someone from complaining by making them feel a sense of duty to voice a complaint or embarrassment at complaining about something trivial. Some of these situational factors may distort the representativeness of complaints in important ways. For example, maybe a good B-MeX measure should put a particular weight on ensuring that wholesalers provide good service to non-household customers that are moving premises, as this may be one occasion when wholesaler interactions are relatively more frequent.
- **Attribution of responsibility.** Consumers are more likely to complain if they believe the other party was at fault. This is what we want for the purposes of B-MeX design, since the B-MeX measure is intended to incentivise wholesalers to improve the services they deliver. In some cases, the complainant's belief that the wholesaler is at fault may not be shared by the wholesaler. However, this would not necessarily mean that a B-MeX measure using complaints data should be over zealous in filtering out complaints that wholesalers consider unjustified. In a competitive setting, wholesalers would risk losing customers if they chose to ignore complaints from customers who think their wholesaler should be doing a better job.
- **Personality traits.** One study suggested that a more prominent role should be given to complaint-related personality variables, such as propensity to complain, rather than the traditional cost-benefit analyses as

driving decisions to complain.³⁴ Complaining generates costs and benefits that are both financial and psychological. Psychological costs and benefits might include stress and a sense of justice being done. Different people may perceive the costs and benefits differently with confident people, for example, expecting the psychological costs to be lower than a nervous person. Absent any evidence that personality traits correlate with particular types of non-household business or are more prevalent in some regions of England than others, the role of personality traits in determining if a complaint is made may not affect the representativeness of complaints. They may reduce concerns that other factors influencing the decision to complain will distort the representativeness of complaints data.

There are two takeaways from these findings that may be relevant when thinking about how complaints data might be used in a B-MeX regime. First, the complaints data may not be representative of what the generality of non-household customers care about. Second, there may be some instances where the volumes of complaints vary across regions for factors outside the control of wholesalers. We think the former is a more material concern.

5.2.3 Potential use of complaints data to help design B-MeX

How complaints data may help

Complaints data indicate areas where customers are dissatisfied. Moreover, they are based on actual behaviour, so avoid problems such as framing bias that may arise when relying on survey responses. The fact that a complaint was made is arguably evidence that the detriment is above a certain threshold, given there will be some costs to making a complaint. This may mean more weight should be given to them when designing a B-MeX scheme, as they allow the focus to be on events causing more serious detriment.

The existing complaints data that CCW monitors, including the summaries of written complaints that retailers receive, should help inform the design of a B-MeX regime. It will also make sense to track complaints data over time, with a view to potentially updating the regime if more recent complaints data suggest problems with the status quo.

The currently available complaints data could provide information about

- the types of activities where customers are dissatisfied with the wholesaler;
- the specific aspect of wholesaler service that was considered unsatisfactory;
- the types of non-household customers making complaints; and
- how promptly and satisfactorily wholesalers deal with complaints.

All of this information is potentially relevant for designing a B-MeX measure. It should help ensure that the B-MeX regime has regard to aspects of wholesaler service where complaints are relatively frequent, and potentially prioritise types of complaints where the evidence suggests wholesalers are slow to adequately resolve the issue.

Complaints data may also be used to identify factors outside the control of individual wholesalers that appear to correlate with customer dissatisfaction, allowing informed decisions to be made about whether and how to control for such factors in the B-MeX regime. For example, if complaints appear to be more common from certain business types or following episodes of flooding, then relative rankings may need to control for differences in the prevalence of certain businesses or flooding in different wholesaler regions. Nonetheless, for reasons set out elsewhere, we advise that the threshold for taking any regional differences between wholesalers should be high.

³⁴ Moshe Davidow and Peter A Dacin (1997). "Understanding and influencing consumer complaint behaviour: improving organizational complaint management" in *Advances in Consumer Research*, Volume 24.

The data could also be monitored to provide an indication of whether non-households' satisfaction with wholesaler service was improving over time, and possible changes in the aspects of wholesaler service that were of most concern. Such findings could then be used to adapt the B-MeX regime. An increase in the volume of complaints may suggest that the current levels of the B-MeX incentives are inadequate and need increasing. This assumes that the complaints have been found to relate to aspects where improved wholesaler service is desired. Using the evidence in this way would implicitly assume that changes in the volume of complaints represents a change in the quality of service provided by wholesalers. Some caution would be appropriate, as it is possible that the change in the complaint volumes reflects, for example, greater awareness about the possibility to complain. It is even possible that complaints are positively correlated with the level of service, with wholesalers' increased focus on customer service meaning that customers find that their complaints are more likely to be resolved by the wholesaler or customers coming to expect higher standards than they were previously willing to tolerate. Better service leading to more complaints leading to a conclusion that service has declined would be a perverse outcome.

Limitations of using complaints data to determine the scope of B-MeX

We would not recommend just relying on the complaints data to determine what should be included in B-MeX. It is likely that complaints data do not fully represent the features that would determine whether non-household business customers would continue using their current wholesaler if they were able to switch wholesaler. Most obviously, there is a risk that complaints data will disproportionately reflect instances of greatest detriment, while omitting many instances where wholesaler service could have been better but was not so unsatisfactory as to prompt the non-household customer to make a complaint.

In a competitive market, firms do not just compete to have fewer dissatisfied customers. They can gain market share by providing a superior service, above and beyond what rivals are offering. In the long run, rivals may imitate such innovations and this higher level of service may become the new normal such that customers would complain if it was not provided. But in the short term, relying on complaints data alone (or designing surveys with a focus on what wholesalers had done badly) risks designing a B-MeX scheme that fails to incentivise innovation in how wholesalers interact with end customers in the business retail market.

It would be sensible to complement the findings from analysis of complaints data with other sources of information, such as survey evidence or monitoring what trade bodies are saying about the service provided by water wholesalers. These alternative approaches are not panaceas, so the complaints data will certainly be useful. For example, a survey-based approach runs the risk of not capturing problems with wholesaler service because the design of the survey did not anticipate the problem being important. The complaints data may serve as a check that important emerging issues are not ignored.

Another risk is that some complaints may entail a conflict between the interests of the complainant and those of other customers. For example, a complaint about how a leakage allowance request was handled may well be correlated with unhappiness that the request for an allowance was rejected or only partially met. The generality of water users in the region may support the wholesaler's leakage policy and not want the wholesaler acquiescing to the request. It is also possible that if they encountered the same scenario, they would not even have sought a leakage allowance, much less complained about how the request was addressed. Just counting complaints without regard to how the generality of customers feel that a wholesaler should deal with such issues runs the risk that the B-MeX regime will incentivise wholesalers to behave in ways that work to the detriment of the generality of users.

Limitations of using complaints data to determine the size of financial incentives

There are problems with using complaint data to infer customer detriment, since an individual complaint does not provide any measure of the value of detriment suffered by the customer. Even if we were able to assign a monetary value of detriment suffered by the complainant, we would need to find some way of scaling this to the entire population. The number of complaints will be a subset of the number of instances where

customers would have valued better service from the wholesaler. But just looking at the complaints data will not provide information on what that relationship is between the detriment suffered by complainants and the generality of customers.

As a starting point, the ratio of complaints about wholesalers relative to complaints about retailers could be used as a guide to the appropriate level of any financial incentives associated with B-MeX in England given the financial incentives facing Welsh water companies in the business retail market. Given the financial incentives governing Welsh water companies in the business retail market cover both wholesaler and retailer activities, it might be reasoned that if 25 per cent of all non-household consumer complaints in the English business retail market relate to wholesaler performance, then the financial incentives facing English wholesalers should be set at one quarter of the level of the financial incentives Welsh water companies face in the business retail market. But such a calculation would be a rough and ready calculation. We would not suggest relying exclusively on complaints data to determine the level of any financial incentives.

5.2.4 Potential use of complaints data as part of B-MeX score

It would also be possible to use the complaints data to calculate the B-MeX scores and determine any financial penalties. The simplest option would be to use the number of complaints as a measure of wholesaler service. Some normalisation of such data would be necessary if using this variable in any B-MeX scheme that featured relative performance, given Thames Water has more than 30 times as many supply points as Sutton and East Surrey Water (504,170 versus 14,013).³⁵

In theory, more of the detail associated with complaints, and not just the volume of complaints, could be used to calculate a B-MeX score. For example, different weights could be given to complaints depending on the type of complaint or how far the complaint was escalated before it was adequately addressed.

Using complaints data should incentivise wholesalers to take actions to avoid complaints being made. If the scoring also distinguished according to how well wholesalers dealt with complaints as they arose, then it may prompt wholesalers to develop better process for handling complaints, which would be a second-best outcome (having no reason to complain would be better).

Many of the limitations discussed above when looking at using complaints data to help with the design of a B-MeX scheme also apply if thinking about using such data in the actual B-MeX metric. For example, the score may not be representative, instead drawing on feedback from customers with a higher propensity to complain and potentially focusing on high-value complaints rather than potentially widespread, minor customer detriment. This concern might be alleviated by using complaints data in conjunction with survey evidence to determine an overall B-MeX score.

Practical problems with using complaints data to determine B-MeX scores

Absolute numbers of complaints by non-households are lower than for households. Whereas there were almost 85,000 written complaints by households about their water company in 2019/20,³⁶ there were less than 15,000 complaints by non-household customers and less than a quarter of these were even partially attributable to wholesalers based on retailer assessment. With 15 equally sized wholesalers, we would expect 200-250 complaints per wholesaler per year. But the wholesalers are very different in size, such that the expected annual number of complaints for the smallest water-only wholesalers could be in the teens. It is possible to imagine a B-MeX relative ranking where a single complaint moves a small water company from being the best to being the worst performing wholesaler if complaint volumes are normalised to permit cross-company comparison, particularly if attempts are made to segment complaints by type.

³⁵ MOSL: "Number of Supply Points by Wholesaler" [\[online\]](#).

³⁶ Discover Water: "Complaints" [\[online\]](#).

Smaller wholesalers may conclude that their measured performance, to the extent that it relies on the volume of complaints, is largely down to noise (did they deal with a customer pre-disposed to complain) rather than underlying levels of performance. In the long run, this randomness should work itself out, but if the relevant staff in the water company conclude that their efforts will only have a marginal impact on their B-MeX score, their personal incentives to improve their firm's performance may be muted.

The accuracy of the complaints data becomes more important if it has financial or reputational implications. This may require additional guidance and an arbiter to consider issues such as whether retailers are collecting and reporting complaints data correctly, whether the complaint is justified, whether decisions to attribute the complaint partly or wholly to the wholesaler are appropriate and, depending on the design of the B-MeX measure, whether the wholesaler service to which the complaint related had been correctly categorised.

Decisions would be needed on the threshold at which a complaint can be disregarded as baseless for the purposes of calculating a B-MeX score. Some complaints may have no merit. We doubt that this will be true of the majority of complaints, but with the potential for just one extra complaint to account for 5 per cent or more of the total complaints for the smallest water-only companies it would not be appropriate to ignore the possibility if there are material financial or reputational consequences associated with each complaint. Wholesalers will understandably want care taken to ensure that unreasonable complaints are excluded from any B-MeX score. This would add to the regulatory burden associated with using complaints data in the B-MeX indicator. The challenge is made greater by the fact that the ultimate objective of the B-MeX regime is better service by the wholesaler. A complaint may make unfair allegations while simultaneously reflecting a genuine dissatisfaction with how the wholesaler interacted with the customer.

Care would be needed to make sure that the outcomes do not just reflect differences that could be attributed to the retailer. Some retailers' customers may be better informed about the possibility of complaining than other retailers' customers. This could mean that financial incentives from the B-MeX scheme reward or penalise wholesalers because of the market share of different retailers in their region, rather than because of differences in how the wholesaler actually interacts with business customers.

Incentive effects with using complaints data to determine B-MeX scores

A B-MeX regime that used complaints data may create incentives for parties to try and game the system. This could distort relative performance rankings based on complaints data, unless all wholesalers were willing and able to game the system equally.

For example, wholesalers may have an incentive to steer non-household customers to register any complaints with them rather than writing to their retailer or CCW. This need not be to the detriment of non-household customers. For example, if the wholesaler immediately acted to rectify the problem, the customer may be better served than had they complained to the retailer or CCW.

More serious would be the possibility that time and effort is spent managing how customers make complaints for the sake of a good B-MeX score, without any corresponding change in the quality of service provided by wholesalers. Massaging the score associated with complaints data becomes the goal. A better outcome would have been for the wholesaler to have proactively addressed such issues so the complaints did not arise. Any time a customer has reason to complain, whether to CCW, the retailer or the wholesaler, that represents a failure of customer relations that the wholesaler would be incentivised to avoid pre-emptively if they were subject to a competitive threat.

The incentive problems need not be confined to wholesalers. For example, it may be that including complaints data in the B-MeX regime will give some retailers incentives to encourage customers to complain (or not complain) about certain things. For example, maybe the retailer perceives that a change in wholesaler service levels as they relate to certain activities will especially help that retailer's business model. A wholesaler receiving bad B-MeX scores because of lots of complaints relating to leakage allowances may be inclined to treat leakage allowances more generously in the future. If this helped a given retailer's business model,

perhaps reducing switching and making it easier to retain existing customers (attractive for retailers with a large market share in a given region), then the retailer may encourage customers to complain about how the wholesaler dealt with their leakage allowance request.

A priori it is not possible to conclude with certainty how serious the problems of perverse incentives will be in practice if complaints data are used to determine a B-MeX score. But assuming that parties respond to incentives – a necessary assumption to justify introducing a B-MeX regime – includes assuming that they may respond in ways that do not align with what the regulator wants to achieve. One way to mitigate these risks would be to reduce the size of the financial penalties and rewards associated with complaints data, possibly by combining the complaints data with other evidence to form an overall B-MeX score. Of course, such a measure also reduces the incentive for firms to attempt to reduce the number of complaints. In other instances, it may require the regulator to engage in ongoing monitoring and being willing to step in and introduce refinements or restrictions to address an unintended consequence. We recognise there is a complaints process at present, but if complaints were to directly influence a B-MeX score, with financial implications for wholesalers, then companies may be more incentivised to game the system than previously. That said, it may be possible to address increased incentives to game via amendments to the existing complaints process. For example, if there was a suspicion that the reported complaints data were lower because wholesalers were steering non-household customers to complain to them and bypass retailers and CCW, it may be that Ofwat or CCW could require wholesalers to start reporting on complaints they receive from non-household customers and incorporate these results into the complaints data used to determine the B-MeX score.

Attaching financial penalties and rewards to the volume of complaints received will certainly not incentivise water companies to encourage complaints that will count in their B-MeX score. Ideally, business customers should have nothing to complain about, but in practice there are always things that the water companies could do better. In this regard, complaints data can be a valuable source of information. Firms in competitive settings are advised to find ways to make sure that customers voice their complaints, so as to help with customer retention, with some reports claiming that over 90 per cent of dissatisfied customers will not complain to the firm but instead switch to a different provider.³⁷

5.3 Recommendations

We recommend using complaints data to help with the design of a B-MeX regime, but that complaints should not be used in the actual scheme. Table 5 below summarises our evaluation. Some of the advantages and disadvantages of using complaints data to determine the B-MeX score could be offset by including other measures, alongside the complaints data, to determine the overall B-MeX score. These are initial recommendation and further exploration and consideration of the use of complaints data during the more detailed design work will be required.

³⁷ Reputation Refinery: “96% of unhappy customers won’t complain to you, but will tell 15 friends” [[online](#)].

Table 5: Evaluation of using complaints data for the purposes of B-MeX (positive points denoted by “+”, negative by “-”)

Evaluation criteria	To help with design of B-MeX	To determine B-MeX score
Customer interest	<ul style="list-style-type: none"> + It identifies issues in the business retail market that cause sufficient dissatisfaction that customers are motivated to complain - Complainants and complaints may not be representative of what the generality of customers care about 	<ul style="list-style-type: none"> + Wholesalers will have incentives to address the things customers complain about, which will tend to be the things that have the biggest impact - Lots of low-level detriment may go unaddressed - No incentive to offer superior service above and beyond what is needed to avoid complaints
Effectiveness	<ul style="list-style-type: none"> + Complaints data already broken down into complaints for which wholesaler wholly, partly or not responsible for. 	<ul style="list-style-type: none"> + Incentivises wholesaler to improve service and avoid complaints - Runs risk that wholesaler and, possibly retailers, will have incentives to concentrate on gaming the complaints process rather than focusing on offering a good service to customers
Cost/practicality	<ul style="list-style-type: none"> + The complaints data already exists - The data provides little guide to what level of detriment a complaint causes (and therefore what financial incentives may be appropriate) - Would need to undertake considerable additional work to complement the complaints data. 	<ul style="list-style-type: none"> + There is already a process in place for collecting complaints data - The work required to make sure that the data are robust and accurate will increase - Relatively few complaints, particularly for the smaller wholesalers

The information in complaints data should be one input into the final design of the B-MeX regime, and should be complemented by other information, some of which will likely have to be collected. Complaint data indicate areas where customers are dissatisfied. Moreover, they are based on actual behaviour, so avoid problems such as framing bias that may arise when relying on survey responses. The fact that a complaint was made is arguably evidence that the detriment is above a certain threshold, given there will be some costs to making a complaint. This may mean more weight should be given to them when designing a B-MeX scheme, as they allow the focus to be on events causing more serious detriment. However, this comes at the expense that they may miss out on large-scale low-level detriment. A firm subject to effective competition would risk losing lots of market share if it only paid attention to customers who had experienced especially bad service.

Once the scheme is up and running, complaints data will continue to be a valuable source of information on monitoring the concerns of non-household customers. If the B-MeX regime is not incorporated into the price reviews, we would recommend that complaints data are reviewed at least once every five years. Changes in the overall volume and type of complaints may help identify possible changes to what is measured in determining the B-MeX score or prompt decisions to vary the financial incentives. If the firms with the best B-MeX scores do not correspond to the firms receiving the fewest complaints (suitably normalised), then that may warrant considering possible changes to the B-MeX measure.

We do not recommend using the complaints data to fully determine the B-MeX score. There are practical problems, including the fact that there are likely to be relatively few written complaints against the smaller water-only companies. There are also potential incentive problems although there may be steps that could be taken to address increased risks of gaming. These concerns may be partially mitigated if the complaints data only represent a part of the final B-MeX score. But if there is separate survey evidence on customer experience gathered from the generality of users, the value of using the complaints data as well to determine a final B-MeX score is reduced. The survey can be designed to give more weight to customer experiences

that are especially bad and therefore might be expected more likely to generate complaints. The main advantage of complaints data would thus be the possibility of capturing aspects of the customer experience that the current survey does not cover. But such complaints would be outliers assuming the survey has been designed appropriately (at worst, there may be occasional lags between certain types of complaints becoming frequent and the B-MeX survey being updated to capture this aspect of wholesaler service).

6 Relative or Absolute Performance

6.1 Introduction

A relative performance measure means that company performance is evaluated relative to that of other companies whilst an absolute performance measure means that company performance is assessed against absolute benchmarks. This chapter considers the arguments for and against the use of these measures of performance in the context of the B-MeX incentive.

In this chapter, we consider the use of both absolute and relative measures of performance for a B-MeX measure. Section 6.2.1 provides a summary of some of the performance measures used by Ofwat and other regulators. Sections 6.2.1 and 6.2.2 then discuss the merits and disadvantages of relative, and absolute measures of performance, respectively. Section 6.2.3 summarises our findings before giving our recommendation in section 6.3.

6.2 Description and analysis of issues

6.2.1 Performance measures used by Ofwat and other regulators

In the PR19 final determinations, Ofwat introduced a series of performance commitments (some with attached ODIs) that were benchmarked against an absolute level of performance.³⁸ This included bespoke performance commitments (and their respective ODIs) focusing on business customer satisfaction for water companies located in Wales.

As part of its 2019 price review Ofwat also introduced incentives focusing on the experience of residential customers (C-MeX) and developer service customers (D-MeX) for the 17 largest water and wastewater companies operating in England and Wales. There were in the form of relative performance commitments. This also means that league tables including the ranking and scores of each company are published by Ofwat on an annual basis.³⁹

In the case of C-MeX, companies receive a score based on the satisfaction ratings given by customers in monthly surveys. Ofwat publishes these scores in annual league tables. Companies can receive rewards in the form of outperformance payments or penalties in the form of underperformance payments. Outperformance payments for a given year can be up to 6 per cent of that year's annual allowed residential retail revenue and underperformance payments can be up to 12 per cent of the residential retail revenue. Rewards for outperforming companies can rise as high as 12 per cent of residential retail revenue if they meet certain requirements: the company is one of the top three performers by C-MeX score; and it is at or above a cross-sector threshold of customer satisfaction performance based on the all-sector upper quartile (ASUQ) of the UK Customer Satisfaction Index (UKCSI); and it has lower than the industry average number of household complaints (per 10,000 connections).⁴⁰

³⁸ Ofwat (2019): "PR19 final determination Delivering outcomes for customers policy appendix" [[online](#)].

³⁹ Ofwat: "Customer and developer services experience" [[online](#)].

⁴⁰ Ofwat: "Customer and developer services experience".

Box 4: Ofgem – Interruption Incentive Scheme (IIS)⁴¹

In the energy sector, Ofgem uses the IIS to ensure the quality of service for customers. Under this scheme, Ofgem gives distribution network operators (DNO's) rewards and penalties annually based on their performance against their targets for the number of customers interrupted per 100 customers (CI) and the number of customer minutes lost (CML). The amount of reward and penalty is based on the responses from a customer survey.

Based on CI performance 1.2 per cent of the DNO's revenue is exposed to rewards and penalties, while 1.8 per cent of the revenue is exposed to rewards and penalties based on CML performance.

The Retailer Measure of Experience (R-MeX), introduced in 2020, allows for a direct comparison between the different wholesalers.⁴² There are currently only reputational incentives associated with getting a good R-Mex score. The measure is based on retailers' reviews of the services provided by wholesalers, and published in the form of league tables so wholesalers understand where they rank relative to other wholesalers.

6.2.1 Relative performance measure

One attraction of a relative performance measure is that it may create similar incentives to those that firms operating in a competitive market might face. In a competitive setting what matters is how a company performs relative to other companies, rather than meeting any absolute levels of performance. A company that satisfies legal requirements or standards set by some monitoring body may nevertheless lose market share if its rivals exceed these levels of performance. It is also possible that if all companies in a sector perform poorly then some of the least poor performers get rewards for being better than a bad bunch.

A relative performance scheme would create exactly the same challenge for wholesalers that faces firms in competitive sectors: offer a better level of service than your rivals. It encourages wholesalers to compete on the service quality to end customers, even though individual customers cannot change wholesaler if they consider other wholesalers offer a better level of service.

A further advantage of relative performance measures is that the informational requirements on regulators are less. They do not have to determine what is achievable and set a target. In the case of an absolute performance measure, given the information asymmetry between regulator and regulated companies there is an inherent risk that regulators set the performance targets above or below the competitive level. Relative measures of performance reduce the problems associated with information asymmetry and do not require the regulator (or market operator) to determine what constitutes an achievable but good level of performance.

Instead, the 'targets' evolve over time. In each period a wholesaler needs to perform better than other wholesalers. This may be an especially attractive feature when setting a multi-year price control, where forecasting what is possible many years into the future may be especially difficult. It removes the risks that the targets are either easily met by all wholesalers or are clearly unobtainable, potentially lowering the incentive for wholesalers to strive to improve service. These risks do not arise with relative measures of performance, as it incentivises companies to improve their services through providing incentives to outperform each other at every stage of the price control period.

Wholesalers' performance need not be affected by common shocks affecting the whole industry. For example, if poor weather across the country prompts business customers to report worse customer experience and therefore a lower B-MeX scores, the best performing wholesaler continue to fare better than the worst

⁴¹ Ofgem: "Quality of Service Incentives" [\[online\]](#).

⁴² Ofwat (2020): "Wholesale Retail Code Change Proposal – Ref CPW084" [\[online\]](#).

performing company, just as would happen in a competitive market. The regime adjusts to average sectoral performance, rewarding firms that are above average and penalising those that do worse than average.

Reporting lags may be considered a problem with a relative performance scheme. Firms will not know how they have fared relative to other firms, and therefore what the financial implications will be of their current offering in the market. This may be perceived as unfair on the wholesalers. It may also affect how much effort wholesalers make to improve service, with no signals to prompt improvements. Firms may settle on aiming to realise a B-MeX score roughly in line with recent industry averages.

A relative performance scheme may be considered to lack ambition. If there is little variation between the firms, it is possible that none of them feel much motivation to improve their offering. This could be addressed by increasing the financial incentives to the top performing wholesaler, or alternatively the regulator may decide to complement the relative performance scheme with an absolute performance target that sets a floor on the minimum level of performance that all firms have to achieve.

Cardinal or ordinal measures

In terms of designing a relative performance measure, consideration should be given to using an ordinal or cardinal approach to relative measures of performance.

An ordinal approach means that it is the relative ranking of each company that matters for the purposes of determining any financial payments associated with the incentive scheme. Under an ordinal approach a company ranked first may face lower incentives to keep improving its performance if it does not get rewarded for the further effort. This could be mitigated by introducing additional financial incentives for firms that improve on their previous scores – the wholesaler effectively competes against other wholesalers and against itself.

In the case of a cardinal approach, the absolute difference in B-MeX scores would affect the level of out- and underperformance payments for each company. Therefore, using a cardinal approach would require forming a view on how much value to attach to different B-MeX scores. This will be especially challenging in the early years when there are no robust and reliable data on company performance. Nevertheless, a possible attraction of using cardinal rankings is that companies for which the evidence suggests relatively little difference in service offerings receive broadly similar financial outcomes. In contrast an ordinal approach may lead to very different consequences for companies ranked third and tenth, even if it was generally considered that there was little to choose between all the water companies bar the top two performers. Of course, if there are meaningful financial consequences for small differences in the B-MeX measure, the incentives for wholesalers to realise improvements will be greater.

For both C-MeX and D-MeX, Ofwat uses a cardinal approach where the financial reward or penalty for a water company is based on its score not rank.

Rewards and penalties under a relative performance scheme

Under a relative performance measure, the regulator can determine the net financial rewards or penalties that wholesalers collectively will receive. This may be important if the central fund of the MPF was used to fund outperformance payments. For example, the rewards could be structured such that it is a zero-sum game: the funds paid in by poorly performing companies exactly offset the outperformance payments made to the wholesalers that performed relatively well compared to their peers.

In the case of a B-MeX incentive implemented through the central fund (as per the current MPF, discussed in Chapter 7), this could mean that the measure is self-sustaining as the amount paid into the fund (as underperformance payments) will equal the amount paid out by the fund (in the form of outperformance companies). For example, the incentive may be calibrated in a way such that wholesalers delivering services to the customers that get a better than average service (e.g. corresponding to approximately half of the

eligible business retail customers in the market) would get an outperformance payment, while wholesalers delivering services to the customers that get a worse than average service would pay into the fund.

However, creating financial incentives for water company shareholders and employees to outperform other wholesalers could be retained if the relative performance scheme was incorporated within the price control framework. The key difference is that it would be the wholesaler's consumers rather than rival wholesalers who funded good performance (through an uplift in allowed revenues) and benefitted financially from poor performance (because the wholesaler's revenue allowance was lower). Locating a B-MeX incentive scheme using relative performance rankings within the price control framework as another ODI would have the attractive property that customer base receiving relatively good service levels pay slightly higher bills while customers of poorly performing wholesalers face slightly lower bills. One practical problem that would need to be addressed with this option is ensuring that it is the wholesalers' business customers whose bills are affected by relatively good or bad B-MeX performance. Because the English wholesalers do not operate in the business retail market, any ODI would have to attach to their wholesale business, which also serves household customers.

For a relative performance measures to work the measure needs to be normalised across companies to ensure that company performance is compared on a like-for-like basis. Assuming that the B-MeX measure depends on survey evidence, thought will need to be given to whether and how to control for differences in the mix of customers surveyed in different regions, differences in the activities undertaken by different wholesalers (including the distinction between water only versus water and wastewater companies), and differences in factors outside the control of wholesalers. If evidence emerges, for example from a pilot B-MeX scheme, that certain customer types or wholesaler activities tend to generate poorer B-MeX scores, then it should be possible to control for this using statistical techniques to adjust raw scores.

Relative measures of performance may be less appropriate when there are material differences between wholesalers that affect their ability to achieve similar B-MeX scores. At this stage, the main risk in this regard seems to be instances where the customer experience depends on both the wholesaler and the retailer. Other possibilities might be that certain events, such as flooding, occur more frequently in some regions than in others and these shape customer perceptions of the wholesaler's performance and consequently affect survey scores, or that regional differences lead to customers placing different weights on what they think is important when thinking about wholesaler performance (e.g. business customers in regions where water scarcity is of greater concern may be less worried about how leakage allowance policies are communicated).

There may be some solutions that permit a relative performance regime to be maintained even where differences in the situations facing wholesalers are identified.

- Creating more than one B-MeX score, whereby companies are ranked separately for the different services provided. In this case, there could be separate financial payments attached to each of the B-MeX scores achieved by the company (e.g. a wholesaler may do well on issues related to supply interruptions and receive an outperformance payment, while ranking relatively low when responding to leakage allowance requests and incur an underperformance payment).
- Creating a composite B-MeX score where the scores received for individual activities are aggregated to give a single B-MeX score for each company which is used to determine the financial payments, but with different weightings used for different wholesalers when determining the composite score. If the weightings were correctly calibrated, the relative performance scheme would reward wholesalers that best addressed the concerns of their customer base given the characteristics of their region. In practice, determining appropriate weightings that everyone has confidence in is unlikely to be achievable.

These solutions would create their own challenges. There will be practical limits to how finely we can define wholesaler interactions. Trying to create lots of separate league tables may be undermined by the fact that for many of those tables only one or two water companies actually have any data to report for the period in question. The precise design of such schemes may materially affect which wholesalers gain or lose the most

from the B-MeX regime, which may result in considerable resources being devoted to trying to influence regulatory decisions on design rather than in serving business customers.

6.2.2 Absolute performance measure

Absolute performance measures can be used to incentivise companies to improve performance through setting challenging targets for the services provided. In particular, these measures can be well suited to circumstances where differences in the activities carried out by companies mean that relative measures are difficult to calibrate objectively. Nonetheless, there are also some limitations associated with absolute measures. In general, we consider that the threshold for adopting an absolute measure of performance should be high.

As noted above, a key challenge associated with setting targets using absolute measures of performance by regulators is asymmetry of information. There may be a concern that the targets ultimately fail to incentivise improved performance. Even if the financial incentives are designed such that a firm is always rewarded (penalised less) for better performance, even when a long way away from the target set, there may be concerns that the outcomes will be perceived as unfair on either consumers or wholesalers. This may, in turn, create reputational damage for the regulator. For example, in the context of the RII0-1 price controls set by Ofgem, Citizens Advice argued that a more robust and systematic package of incentives for network companies could have saved £1.1bn for customers. In Citizens Advice's view this tougher package of incentives could be achieved by matching the rewards paid to the best performers by penalties paid by the poorest ones (an example of the zero-sum game discussed above), instead of using an incentive system that was asymmetric in companies' favour.⁴³

A possible way in which the regulator can ensure that the targets are challenging for companies is to have a periodic review of the effectiveness of the benchmark. A potential problem with too frequent adjustment or recalibration of targets (e.g. occurring every year using regulatory discretion instead of every five years corresponding to price control review timelines) is that this creates uncertainty for regulated companies which in turn could affect their business plan proposals and commitments in terms of improvements to the services delivered to customers.

6.2.3 Summary of relative and absolute measures of performance

Table 6 below summarises and assesses the relative merits and disadvantages of using relative and absolute measures of performance against the evaluation criteria.

⁴³ Citizens Advice (2017) "Energy Consumers' Missing Billions - The profits gifted to energy networks", [[online](#)].

Table 6: Evaluation of relative and absolute measures of performance against criteria (positive points denoted by “+”, negative by “-”)

Evaluation criteria	Relative performance	Absolute performance
Customer interest	<ul style="list-style-type: none"> - Does not guarantee a minimum level of service is achieved across all regions – though this can be mitigated by having a minimum performance requirement for the water companies. - May not be possible to design such a scheme to ensure each wholesaler’s incentives align with the preferences of that wholesaler’s customers. 	<ul style="list-style-type: none"> - A regulator may not set sufficiently challenging targets. + Companies can provide similar / same level of service over time. + Can reflect a base level of what customers expect and need.
Effectiveness	<ul style="list-style-type: none"> + Can be used to mimic competition and create pressure similar to those companies would face in a competitive market. + Through the MPF the incentive scheme can be self-sustaining i.e. amount paid into the fund equals amount paid out of the fund. Effectiveness will also depend on whether a cardinal or ordinal approach is taken. + Creates rolling targets for companies. - Possibility that different wholesalers’ ranking depends on exogenous factors, including differences in the market share of retailers in different regions 	<ul style="list-style-type: none"> - Due to information asymmetry, targets may be too easy/difficult for companies to achieve. + Can be used to improve industry performance over time, forcing the industry to strive to achieve performance levels they might otherwise have considered unobtainable. + Allows for bespoke targets for each wholesaler that responds to the particular situation facing that wholesaler. - Uncertainty about the regulatory targets that might be set in the future may distort behavior, - Exogenous factors may affect companies’ performance and in doing so result in rewards and penalties accruing that do not relate to wholesaler effort.
Cost/practicality	<ul style="list-style-type: none"> - Measure will need to be normalised to ensure a like-for-like comparison which can be difficult in practice. - Exogenous factors may make it very difficult to design - Requires confidence that the data from all regions is collected on a like-for-like basis. 	<ul style="list-style-type: none"> - Performance level may be difficult to determine in absence of (reliable) data and due to issues around information asymmetry. - Targets to be updated from time to time which in turn can create a design issue for the regulator.

6.3 Recommendations

We recommend that the default should be for the B-MeX incentive to be based on a relative measure of performance. The threshold for accepting deviations from this and adopting an absolute measure of performance should be high. A relative measure of performance would introduce benchmark competition between wholesalers and would avoid Ofwat having to overcome the informational asymmetries and identify a suitable absolute level of performance.

If there are some activities where an absolute measure is deemed appropriate (e.g. if there are differences between the regions affecting business customer experience that are outside the wholesaler’s control, supported by compelling evidence), we would still recommend retaining a relative measure of performance for other activities to be covered by the B-MeX incentive.

7 Regulatory Instrument for B-MeX Incentive

7.1 Introduction

This chapter investigates the regulatory instrument that may be used for a financial B-MeX incentive. The three alternative options considered include:

- financial payments through a central fund (similar to the current MPF approach);
- the price control framework; and
- offering direct compensation to customers (e.g. through guaranteed standards of performance).

Section 7.2.1 explores the use of a central fund (as currently used for collecting the payments associated with MPS and OPS), section 7.2.2 considers the price control framework as a potential regulatory instrument, and section 7.2.3 investigates whether direct compensation may be offered to end customers as part of the B-MeX incentive. Section 7.2.4 summarises the merits and disadvantages of these alternative options and section 7.3 includes some initial recommendations.

7.2 Description and analysis of issues

7.2.1 Use of a central fund (as per the current Market Performance Framework)

This section considers whether and how a central fund, as currently used within the MPF, may be used to implement and administer a financial B-MeX incentive scheme, where the central fund receives any financial penalties wholesalers make for underperformance. As part of the options explored, we also consider the possibility of using the central fund to fund financial rewards for outperformance or using it to compensate end customers for poor wholesaler performance.

Use of central fund within the MPF

The MPF involves financial penalties being paid into a central fund by wholesalers and retailers associated with poor levels of performance against the Market and Operational Performance Standards (MPS and OPS), administered by MOSL. OPS and MPS focus on wholesaler-retailer interactions and do not reflect the experiences of end business customers. The framework does not include explicit financial rewards for outperformance, and initially the fund redistributed the charges collected from trading parties at the end of the year back to companies based on their market share.

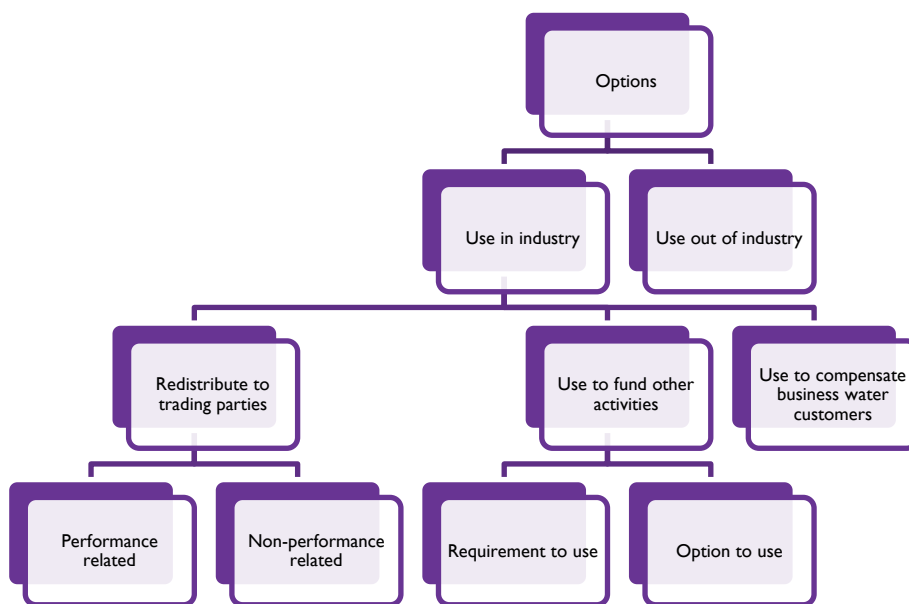
Previous work commissioned by MOSL and the Market Performance Committee (MPC) has looked into the treatment of MPS and OPS charges and considered a range of options (summarised in Figure 1 below) on how the charges collected may be used within and outside the industry.⁴⁴ In particular, the report considered the initial redistribution model used and proposed alternative models in terms of their properties to incentivise companies to improve performance, including examining the possibility of designing a model that does not dilute incentives.

The report shortlisted two options that might be used to improve the services provided to business customers. One option involves redistributing all of the charges paid into the central fund through a

⁴⁴ Economic Insight (2019): “Options for the use of MPS and OPS charges” [[online](#)].

mechanism where companies only receive a share of the charges paid by others, but do not receive a share of the charges the company itself has paid. The second option involves the default arrangement of redistributing 100 per cent of the charges based on the method described above, but allowing for the possibility of using the charges to fund other activities through a pre-defined decision-making process. The report concluded that this second option may be preferred as some activities could be funded through the charges.

Figure 1: Options for the use of MPS and OPS charges



Source: Economic Insight (2019): “Options for the use of MPS and OPS charges”, p.8 [\[online\]](#).

Since the publication of the report, Ofwat implemented a change to the redistribution model in January 2020 to address the weak incentives created by the previous framework. Concerns around the previous model included that the redistribution of the penalty payments paid into the central fund diluted the incentives provided to companies to improve service provision to end customers and undermined the efficacy of financial penalties. This was believed to be more of a concern for the largest companies who – under previous arrangements – may have received back up to 10 per cent of any penalties paid. As a result of the change implemented, under the current MPF arrangements companies may only receive a share of the charges paid by other trading parties (and none of the penalties the company itself has paid).⁴⁵

Governance arrangements within the industry will also be relevant in assessing the implications of using a central fund. As the industry regulator, Ofwat sets wholesale price controls for water and wastewater companies in England and Wales, as well as household and non-household retail controls in Wales every five years. This includes determining the level of water and wastewater services customers receive which are fixed for the price control period. In the case of the MPF, changes to some of the market codes (including the Market Arrangements Code, which is where the MPF is set out) may be proposed by the industry with Ofwat making the final decision whether changes should be implemented. Further, changes could also be raised by Ofwat. There is flexibility for both industry and Ofwat to adjust the MPF, for example in response to any changes in market dynamics.

Potential options to use a central fund for B-MeX incentive

⁴⁵ Ofwat (2019): “Wholesale Retail Code Change Proposal – Ref CPM018” [\[online\]](#).

The current central fund involves penalties paid by wholesalers and retailers which may be used in a number of ways. In this section we first briefly consider four options, ranging from redistributing the fund back to companies to taking the penalties paid outside the water sector. Each of these four options involve underperformance payments only from wholesalers providing poor levels of service to end customers and do not explicitly reward companies that improve performance and end customer experience.

- Conceptually, the central fund could treat any penalties associated with a B-MeX incentive similarly to the way in which penalties for OPS and MPS are collected and used. Therefore, one option for the fund would be to **redistribute the money paid into the central fund to water companies** (and their shareholders). Therefore, while this option does not formally involve outperformance payments, companies providing good levels of service may receive some of the money paid into the fund by poorly performing companies.
- A second option would be to **use the penalties paid into the central fund to compensate business retail customers**. Attempting to target payments to affected customers would be challenging. These include issues around whether the payments would be directly administered by wholesalers or whether these would go through retailers, difficulties in identifying the customers affected by poor levels of performance, and determining the compensation due for different instances on the level of service provided by the wholesaler.
- A third option would be to **use the penalties paid by wholesalers to fund other (not business-as-usual) activities**, e.g. commissioned by Ofwat or MOSL. Business customers would benefit since there would be incentives on wholesalers to improve performance and even when they fail there should be some customer benefit accruing through the additional activities funded by the payments.
- A final option would be to **use the money paid into the central fund for other purposes**, such as donating it to Treasury. This option fares poorly in terms of the customer interest. Should the incentives for better wholesaler performance not have worked, the customer receives no compensation as a water customer.

Therefore, in addition to the four options discussed above involving underperformance payments only, a further option to consider is **providing both penalties for underperformance and (explicit) financial rewards for outperformance**. One concern with including financial rewards for good performance is that the fund may go into debt. It would be necessary to calibrate the payments in the B-MeX incentive to avoid such an outcome. This would be more of a concern for an absolute-performance regime. A relative performance regime could be designed to have zero-sum outcomes.

The possibility of including both under- and outperformance payments for the incentive through the central fund also raises the question of having a separate fund for the B-MeX incentive only, or whether these payments would be combined with the penalties paid by parties associated with the MPS and OPS. In the case of using a central fund that combines the charges for MPS and OPS and the B-MeX under- and outperformance payments, theoretically it may be possible to use the wholesaler penalties associated with MPS and OPS to fund any outperformance payments for B-MeX that are not covered by the underperformance penalties paid by companies for the same incentive. Nonetheless, designing and calibrating such an incentive scheme would be a challenging and complex exercise, and may not be consistent with the governance arrangements of the MPF itself.

A related consideration is that all current incentives provided to companies through the MPF involve penalties only, therefore the use of a central fund for the B-MeX incentive may be more consistent with a penalty-only scheme. This would, for example, be consistent with the approach taken by Ofwat to some performance commitments as part of PR19 such as the Compliance Risk Index (CRI) which targets full compliance with statutory obligations through an underperformance-only ODI attached.⁴⁶ Nonetheless, design and calibration

⁴⁶ Ofwat (2019): "PR19 Final determinations Delivering outcomes for customers policy appendix" [[online](#)].

considerations aside, there do not appear to be any particular reasons as to why the central fund may not support outperformance payments to wholesalers.

7.2.2 Use of price control framework

An alternative option to using a central fund for the payments associated with a financial B-MeX incentive would be to introduce the business customer satisfaction measures within the price control framework. This also means that wholesalers would have a duty to report progress against the measures specified as part of the B-MeX incentive, and Ofwat would apply the relevant payments (under- and/ or outperformance payments) to individual wholesalers at the end of the price control period, or in-period through the usual regulatory reconciliation mechanisms.

An advantage of using the price control framework for the B-MeX incentive would be that some of the design and calibration issues with outperformance payments under the MPF described above could be avoided. For example, in the case of using absolute measures of performance wholesalers' allowed revenue could be adjusted up or down based on actual levels of performance and any out-and underperformance payments attached to these. This means that if all companies outperform certain absolute benchmarks set around end customer experience, all could receive outperformance payments under the B-MeX incentive through the price control arrangements. Such an outcome may not be possible if outperformance payments were paid from an MPF central fund, as there may be insufficient funds.

Nonetheless, the price control route would mean that once the measure is set, there would also be limited opportunities (if any) to apply any changes to the design or level of incentive for the duration of the price control.

A further consideration with using the price control framework is the absence of a business retail price control to attach the incentive to (C-MeX underperformance payments relate to the residential retail price control). The obvious solution would be to attach any penalties to the wholesale price controls instead. An advantage of using the wholesale price controls would be that customers receiving lower levels of service would be facing lower bills.⁴⁷ Nonetheless, the wholesale controls cover services provided to both household and non-household customers, which in principle could mean that low levels of services provided to business customers imply lower bills for residential customers and vice versa. As things currently stand the only protection against such outcomes is that wholesalers, in complying with the tariff principles, might take care to ensure that any costs or revenues associated with a B-MeX ODI are allocated to the non-household wholesale tariffs when those tariffs are determined. A further possible concern is that a B-MeX ODI attached to the wholesaler's network business will represent a much smaller percentage of revenues than the corresponding C-Mex ODI attached to the residential retail price cap, even if the total sums at stake were designed to be equal. It is possible, drawing on ideas from behavioural economics, that this presentational detail will have implications for how motivated water company managers are to improve a poor B-MeX score.

7.2.3 Direct compensation to customers

A third, different option would be to offer direct compensation to customers affected by the poor levels of performance. This option may be implemented inside both the MPF and price control framework, for example through setting guaranteed minimum standards of service (based on some absolute level of performance). Within the water sector the guaranteed standards scheme (GSS) requires water and wastewater companies to make a specified payment to the affected customers who have not received the guaranteed minimum standards of service. The GSS includes service provision around making and keeping appointments, low water

⁴⁷ This is because charges are required to be cost-reflective, and bad debt adjustments to wholesale charges could only be recovered from non-household customers.

pressure incidences, incorrect notice of planned supply interruptions, etc. and specifies the minimum level of payment for companies both in the case of residential and business customers. Companies may voluntarily choose to increase these payments.⁴⁸

Box 5: Delay Repay Scheme in the rail sector

In the rail sector the Delay Repay (DR) Scheme⁴⁹ provides direct compensation to customers for delayed train journeys. DR15 provides compensation for passengers whose train was delayed between 15-29 minutes, while DR30 provides compensation to those customers who experienced delays of more than 30 minutes. Under DR30 passengers experiencing delays between 30 and 59 minutes are entitled to 50 per cent of the cost of a single ticket, and those experiencing a delay of 60 minutes or more are entitled to a 100 per cent refund of a single ticket. If the delay is 120 minutes or more, affected passengers can claim the cost of a return journey (if applicable).

Offering direct compensation to customers may be more advantageous compared with a central fund in terms of promoting customer interest, as compensation would be directly paid to customers who have experienced some form of detriment in a timely manner. Payments may be automatic or triggered by a claim submitted by affected customers.

An issue with direct compensation is that any such measure would need to be based on objective and measurable indicators of performance rather than survey responses gathered from (potentially affected) customers. Offering compensation to customers based on, for example, subjective survey responses would lead to perverse incentives encouraging survey respondents to misrepresent company performance so that they can receive financial compensation. This is particularly relevant in the context of customer satisfaction measures, including B-MeX which are arguably intended to capture the aspects of service provision which are qualitative in nature.

7.2.4 Summary of options around the regulatory instrument for B-MeX incentive

Table 7 below summarises the relative merits and disadvantages associated with the three different options around the regulatory instrument that could be used for the B-MeX incentive.

⁴⁸ Ofwat: “Standards of service” [[online](#)].

⁴⁹ Network Rail: “Technical overview: Payments relating to disruption” [[online](#)].

Table 7: Evaluation of options around using a central fund for B-MeX incentive against criteria (positive points denoted by “+”, negative by “-”)

Evaluation criteria	Central Fund (as per current MPF)	Within price control framework	Direct compensation to customers
Customer interest	- Customer affected by poor levels of performance are unlikely to receive compensation, such as lower bills	+ Customers receiving lower levels of service would be facing lower bills - Due to the wholesale control, could be difficult to ensure that only business customers bills are varied in response to B-MeX outturns	+ Ensures that customers affected by poor levels of performance get compensation
Effectiveness	- May be difficult to reward out-performance (unless relative measure is used) - Incentives to improve performance may be diluted (e.g. if charges are distributed back to wholesalers)	- Limited opportunities (if any) to change the design or level of incentive for the duration of the price control period - Potentially inconsistent with R-Mex and MPF developments + Could be used to support both out- and underperformance payments	+ Wholesalers may face additional payments to customers through direct compensation in addition to any price caps
Cost/practicality			- Customer survey unlikely to work as requires objective and measurable indicators of performance

In addition, any decisions regarding the other policy issues considered as part of the B-MeX incentive in this report (including the use of relative measures of performance, and the activities covered by the measure) will also affect the feasibility and suitability of different regulatory instruments for the B-MeX incentive.

7.3 Recommendations

This chapter has explored three different regulatory instruments in the context of the B-MeX incentive, including the use of central fund (as per the current MPF), the price control framework, and offering direct compensation to customers.

Offering direct compensation to customers experiencing poor levels of service (e.g. using guaranteed standards of performance) could be used promoting customer interest, as compensation would be directly paid to those who have experienced some form of detriment as a result of wholesalers' performance. Nonetheless, any such measure would need to be based on objective and measurable indicators of performance which would be challenging to define if B-MeX focused on the qualitative aspects of the services provided by wholesalers.

In principle both the central fund (as per the current MPF) and the price control framework could be used to incentivise improved wholesaler performance through a financial B-MeX incentive.

- In the case of using the central fund, this may include both penalties for underperformance and financial rewards for outperformance, although calibrating such a scheme would be a challenging and complex exercise, suggesting that a zero-sum scheme based on relative measures of experience may be more appropriate.
- Similarly, the price control framework could also be used for both under- and outperformance payments associated with the B-MeX incentive, although once the measure is set, there would be limited

opportunities (if any) to apply any changes to the design or level of incentive for the duration of the price control period.

Taken together, further evidence and analysis is required before a decision around the regulatory instrument for implementing and administering a financial B-MeX incentive can be made.

8 Size of Financial Incentive

8.1 Introduction

Options and issues around the size of any financial incentives for a B-MeX scheme will require careful thought. The measure needs to provide incentives for wholesalers to provide better service, but not over-incentivise companies to put a disproportionate emphasis on qualitative aspects of the business customer experience at the expense of other service and operational requirements, which customers may value more. As discussed in Chapter 6, relative measures of performance are well suited to mimic the pressures companies would face in a competitive market, therefore any rewards or penalties attached to the incentive should also seek to replicate similar pressures.

This chapter investigates the size of a financial B-MeX incentive through looking at the size of financial incentives attached to the water sector and beyond in section 8.2.1 and the interactions between B-MeX and other existing incentives faced by wholesalers in section 8.2.2. Some other aspects of financial incentive design are briefly considered in section 8.2.3. Section 8.3 gives our recommendation around the size of the financial incentive for the B-MeX scheme.

8.2 Description and analysis of the issue

8.2.1 Size of customer experience measures in the water sector

The current measures of customer experience that are used to incentivise water companies are obvious reference points to consider when designing a B-MeX incentive. There are three such measures of customer experience already in use in the water sector. First, Ofwat introduced a measure of business customer experience for the two companies wholly or mainly located in Wales as part of PR19 (Dŵr Cymru and Hafren Dyfrdwy). Secondly, as part of PR19 Ofwat replaced the service incentive mechanism (SIM) with C-MeX, a customer measure of experience. Finally, in PR19 Ofwat also introduced D-MeX, a measure of experience for developer service customers. Given the similarities between C-MeX and D-MeX in terms of incentive mechanism design, these measures are discussed together in this chapter.

At a minimum, these measures provide possible lessons on the types of issues that need to be addressed and how this could be done. It will also make sense to benchmark B-MeX proposals against these existing regimes to check that the proposals are proportionate given other incentives in place and that the different regimes exhibit some regulatory consistency. Where inconsistencies are identified, it is possible that the appropriate response may be to refine the other schemes (over time) rather than revise the B-MeX plans.

In the case of Wales, both companies saw the introduction of a business customer satisfaction measure in PR19.⁵⁰ These measures are based on a customer survey where a sample of all non-household customers each quarter are asked to rate the services provided by the companies on a scale of 1 to 5. The outcomes are then averaged to give a final score for the wholesaler and, dependent on the benchmarks set the company may receive an out- or underperformance payment in light of the final score achieved.⁵¹ With regards to the relative size of these measures, the financial payments attached to these bespoke performance commitments

⁵⁰ These were introduced as bespoke performance commitments for Dŵr Cymru and Hafren Dyfrdwy.

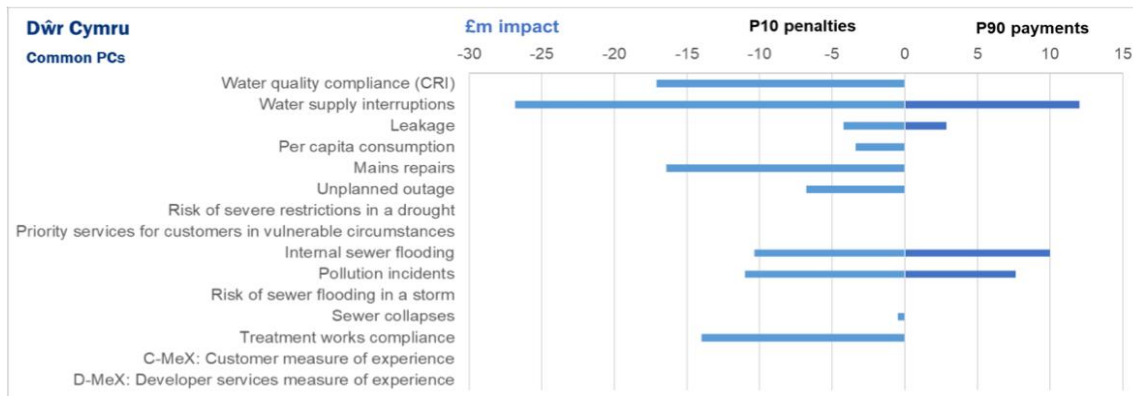
⁵¹ Ofwat (2019) "PR19 final determinations: Dŵr Cymru – Outcomes performance commitment appendix", p.75-77 [[online](#)].

are typically smaller than the payments attached to other common (or bespoke) performance commitments focusing on the operational aspects of service provision such as water supply interruptions, pollution incidents or internal sewer flooding. Figure 2 and Figure 3 below present an indication of the financial value of Dŵr Cymru's ODIs⁵² including the business customer satisfaction measure. In terms of the return on regulatory equity (RoRE), the financial value of the overall ODI package for the company at the final determination corresponds to -1.03 per cent and +0.61 per cent of the 5-year regulatory equity at the upper and lower extreme levels of performance.⁵³

⁵² The same is also true for Hafren Dyfrdwy although figures are not presented in this report.

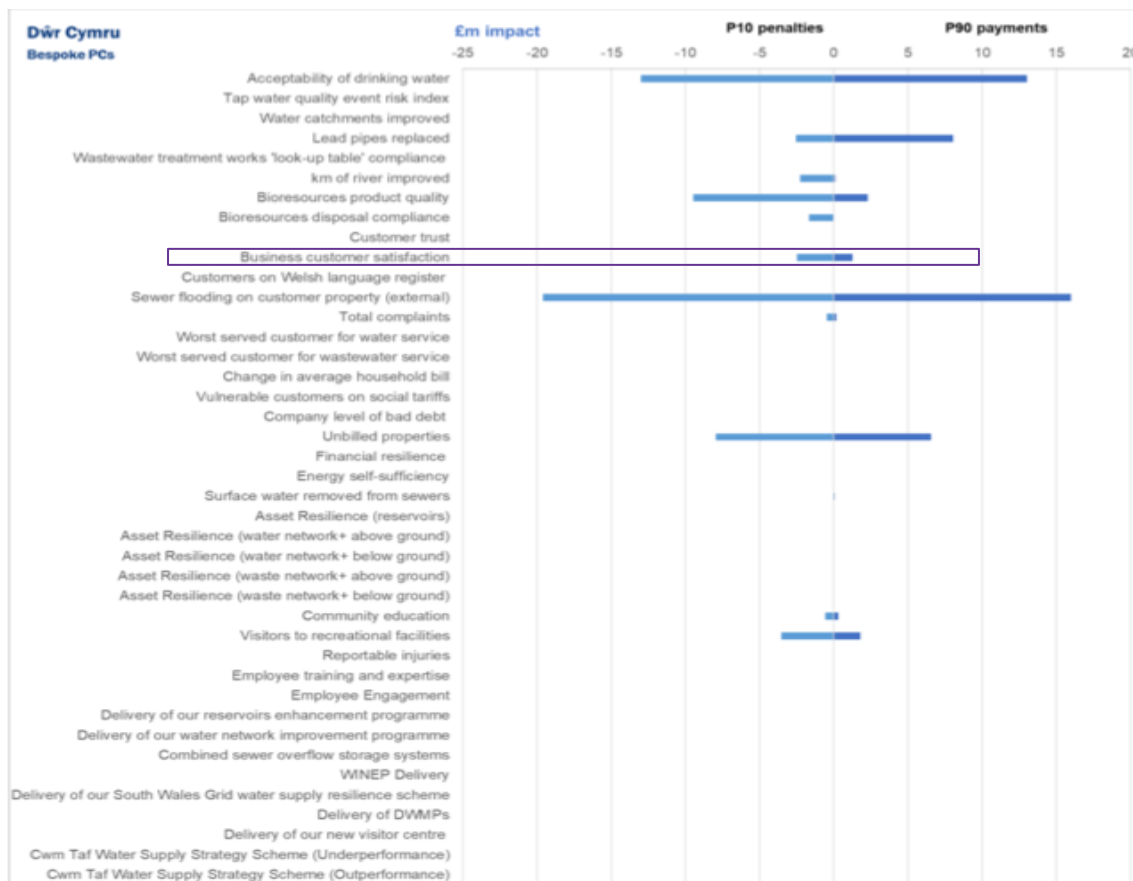
⁵³ Ofwat (2019) "PR19 final determinations: Dŵr Cymru final determination", p.28-29 [[online](#)].

Figure 2: Projected P10 underperformance payments and P90 outperformance payments⁵⁴ for common performance commitments over 2020-25 (£ million)⁵⁵



Source: PR19 final determinations: Dŵr Cymru – Outcomes performance commitment appendix” [online].

Figure 3: Projected P10 underperformance payments and P90 outperformance payments for bespoke performance commitments over 2020-25 (£ million)



Source: PR19 final determinations: Dŵr Cymru – Outcomes performance commitment appendix” [online].

Both C-Mex and D-Mex measures are primarily survey-based incentive mechanisms and as such measure the feedback provided by customers regarding the services provided by water companies. The out- and underperformance payments attached to these measures correspond to between +6 per cent and -12 per

⁵⁴ The P10 is defined as “the performance threshold at which there is only a 10% chance of outturn performance being worse” while the P90 refers to “s the performance threshold at which there is only a 10% chance of outturn performance being better”. These indicate how much Dŵr Cymru would need to return to customers if it underperformed at the P10 level or how much it would gain if it outperformed at the P90 level.

cent of the annual allowed residential retail revenue of each company.⁵⁶ In the case of C-MeX, the outperformance payments could increase to up to 12 per cent of the annual residential retail revenue if:

- the company ranks in the top three performers by C-MeX score; and
- based on the all-sector upper quartile of the UK Customer Satisfaction Index, the company ranks at or above the cross-sector threshold; and
- compared to the industry, it has lower average number of household complaints per 10,000 connections.

8.2.2 Interactions with other incentives

The size of other incentives faced by wholesalers – including the ODIs attached to performance commitments through PR19 – will have implications for the size of any financial incentives for the B-MeX scheme.

Activities relating to water and wastewater operations and maintenance such as water supply interruptions, leakage, sewer flooding, etc, are already captured by the PR19 performance commitments (and the ODIs attached), and typically have significant financial payments attached to them. Similarly, the payments related to C-MeX and D-MeX also represent a sizeable percentage of annual revenue for companies.

The size of existing incentives will also have implications for the B-MeX incentive. First, a B-MeX incentive should not seek to incentivise companies along dimensions that are already covered by existing schemes as this would lead to double counting. However, this does not exclude the possibility of a B-MeX incentive focusing on the qualitative, intangible aspects of service provision complementing any existing, more specific quantitative measures around service provision. For example, where wholesalers are incentivised on the operational aspects related to supply interruptions and leakage through the performance commitment and ODIs set in PR19, the qualitative aspects of service provision (e.g. how promptly the issue was addressed and how workers engage with the customer when working on-site) could still be captured through the B-MeX incentive.

Further, depending on the relative size of a B-MeX incentive and other incentives focusing on activities regarding water and wastewater operations and maintenance, companies may decide to focus more or less on the services provided to end customers just as they would in other competitive market settings. If a wholesaler does not show due consideration to the customer when it is on site repairing a meter, this could lead to a customer deciding to switch in a competitive market. But this would also depend on the value customers place on the different services provided by the wholesaler, e.g. what would a customer be willing to pay to avoid a scenario in which the wholesaler does not show due consideration relative to what it would be willing to pay to avoid operational issues such as supply interruptions or water quality events (if these aspects of performance also differed between wholesalers).

Nonetheless, disproportionately large incentives attached to customer satisfaction could lead to wholesalers prioritising customers experience and other ‘qualitative aspects’ of service provision over operational issues. Taken to the extreme, this could mean incentivising companies to put less emphasis on avoiding service failures (e.g. sewer flooding incidents) in the hopes of mitigating the financial consequences of these incidents through the payments received associated with customer satisfaction measures.

The interaction between wholesaler and retailers, and the incentives wholesalers face with regards to the services provided to retailers through the R-MeX will also have implications for the design of the B-MeX incentive. This includes both the type of the incentive (financial or reputational) and the magnitude of payments attached to each incentive. In terms of the relative sizes, should B-MeX be disproportionately larger

⁵⁵ P10 underperformance payments and P90 outperformance payments for C-MeX and D-MeX, which are relative incentives, are not included in the figure.

⁵⁶ Ofwat (2019): “Customer measure of experience (C-Mex) and developer services measure of experience (D-Mex) policy appendix”, p.19 [[online](#)].

than R-MeX, it could imply that wholesalers put more emphasis on interactions with business customers at the expense of retailer- wholesaler relationships. Given that the experience of retailers also indirectly affects the services provided to end customers this may lead to worse outcomes for business customers. Further, wholesalers may also start improving their relationship with end customers and develop better brand awareness among them.⁵⁷ This could be done through offering more value-added services to end customers that otherwise may have been provided by retailers or interact more with businesses customers without the involvement of retailers, possibly at the dissatisfaction of the latter.

8.2.3 Other aspects of financial incentive design

In addition to the size of the financial incentive and the interaction with other existing incentives, there are further aspects of financial incentive design that will need to be considered when designing the B-MeX incentive. This section provides a descriptive list of these dimensions without considering the detailed design issues associated with these aspects.

In terms of the size of the financial incentive discussed above, this will relate directly to the scope of the activities covered by the B-MeX measure and customers' willingness to pay for improvements in the services provided.

Other aspects of financial incentive design to consider for the B-MeX scheme include:

- **Type of financial incentive:** the payments attached to the incentive could involve underperformance payments only where only those companies performing worse than a specified level (based on absolute or relative measure of performance) would be facing a penalty, or these could involve both underperformance and outperformance payment where those companies performing better than the specified level would receive rewards for the level of services provided.
- **Symmetric or asymmetric incentives:** if the incentive involves both under- and outperformance payments, these could be symmetric (e.g. where in the case of a relative measure of performance the best performing companies would be facing the same magnitude of payments in terms of the rewards received that the worst performing company in terms of the penalties paid) or asymmetric (e.g. where in the case of relative measures the worst performing companies faces larger penalties than the rewards received by the best performing companies in terms of order of magnitude).
- **Higher outperformance payments:** the scheme may involve higher outperformance payments subject to the companies reaching some additional targets or threshold specified in terms of the services provided.
- **Structure of the incentive:** the payments attached to the incentive may be linear or non-linear e.g. depending on the company's ranking in the case of a relative performance measure.
- **Caps and collars:** caps and collars may be used to limit the overall out- and underperformance payments faced by companies.
- **Derivation of an appropriate unit incentive rate:** for many of the PR19 performance commitments water companies used a range of willingness-to-pay (WTP) studies to understand customer preferences and valuations, predominantly focusing on a household setting. These studies aimed at evaluating how much consumers would be willing to pay to receive better services such as cleaner water or higher water pressure. In the case of B-MeX WTP method may be less relevant to draw on since the focus is on customer satisfaction rather than different levels of service provision.

Each of these aspects would need to be considered further before a decision can be made.

⁵⁷ This may especially be the case if the B-MeX incentive is based on based on a general customer survey rather than focussed on specific interactions with the wholesaler.

8.3 Recommendations

Taken together, our recommendation is to introduce B-MeX as a reputational incentive (i.e. with no financial payments attached) until the design issues around the financial incentive discussed above are further explored.

For example, a pilot or shadow period could be used to further explore the size of the financial incentive to ensure that it is significant enough to incentivise wholesalers to provide better services to end customers, but does not over-incentivise companies to put a disproportionate emphasis on qualitative aspects of the business customer experience. The shadow year could also be used to explore customer preferences and valuations around the different services provided, including whether customers would support any outperformance payments attached to the incentive or whether the B-MeX scheme should be introduced as an underperformance only incentive.

Appendix 1: Lessons Learnt from the Water and Other Sectors

Scottish business retail market

The Water Services (Scotland) Act of 2005 created the framework for retail competition for both water and sewerage services in Scotland. In 2006, the Scottish Water retail business was legally separated into a separate entity known as *Business Stream* (although it remained in public ownership). Further, a Central Market Authority was created and market *Operational Codes* were developed to support the opening of the retail market in April 2008 for around 100,000 business customers in Scotland. The market is designed to be as transparent as possible and to minimise transaction costs. The market is therefore built on the basis of regulated, rather than negotiated access, and governed by a set of legally binding market codes.⁵⁸

Non-Household Customer Experience

The non-household Customer Experience Measure (nhCEM) was first introduced by the Water Industry Commission for Scotland during the Strategic Review covering the period 2015-21. The nhCEM is a similar measure to the household Customer Experience Measure (hCEM) however focusing specifically on non-household customers. The household Customer Experience Measure combines qualitative and quantitative components to monitor the domestic customer's experience with Scottish Water, including the views of domestic customers who have had an issue but did not contact Scottish Water about it.⁵⁹

Scottish Water developed, and in 2017/18 started using, the separate **non-household Customer Experience Measure** (nhCEM), aiming to support the company in improving the service and experience it provides to non-household customers. This new measure covers a range of actors from the sector including licensed providers, developers and business customers. The introduction of the measure also meant that for the first time Scottish Water collected feedback from business customers who have contacted the company directly, with a view to understand the different levels of customer expectations, and how Scottish Water can improve customers' experience in dealing with the company.⁶⁰

Since 2017 Scottish Water worked to deliver improvements in the stability of the measure. In addition, in 2017/18 Scottish Water introduced a new survey to better understand their experience of property developer community and establish actionable insights. Since 2018/19 the nhCEM is one of Scottish Water's key business performance targets aimed at improving customer experience for both non-household and developer customers.⁶¹

By 2019/20, the non-household Customer Experience Measure (nhCEM) improved to 85.19 (from 81.74 in the previous year). There has also been an increase in licensed provider satisfaction levels from 90.82 per cent to 95.44 per cent, as well as in Business End User satisfaction levels which has improved from 87.64 per cent to 90.08 per cent during the same period.⁶²

⁵⁸ Welsh Government (2014): "Delivering results for water customers in Wales" [\[online\]](#).

⁵⁹ ERRRA (2015): "The Role of Customers in the Strategic Review of Charges Process in the Water Industry in Scotland" [\[online\]](#).

⁶⁰ Scottish Water (2017): "Scottish Water WIC Annual Return Commentary June 2017" [\[online\]](#).

⁶¹ Scottish Water Annual Report 2017/18 [\[online\]](#).

⁶² Scottish Water Annual Report 2019/20 [\[online\]](#).

Rail

Schedule 8

Schedule 8 is a performance regime in the rail sector which compensates train operators for delays caused by Network Rail as well as by other train operators. It intends to promote both punctuality and reliability for Network Rail and the operators through financial payments, including penalties and rewards based on outturn performance relative to benchmarked levels of performance.

On Schedule 8, any lateness or cancellations incurred by operators' actions implies that these operators will be held financially accountable. However, for any other delays or cancellations it will hold Network Rail accountable even if it did not cause them (e.g. weather conditions).

Benchmarks under Schedule 8 are different for Network Rail and train operators and were last recalibrated as part of PR18. These are based on average lateness in a four-week period. If during that period train operators or Network Rail perform at their benchmark, no payments will be required. In the case where train operators perform worse, they will have to compensate Network Rail while if they exceed their benchmark they will receive compensation.⁶³

The Network Rail payment rates to (and from) train operators are based on "the assessed marginal revenue effect (MRE) of poor performance to that service group".⁶⁴ The impact of poor performance on different passenger types (e.g. commuters or leisure travellers) was also taken into account as part of the PR18 recalibration of payment rates.

Passenger train operator payment rates (i.e. the rates paid by train operators for the delays they cause) are also set at the service group level and are designed to reflect the impact of one train operator's performance on another (which may go beyond the train operators whose services run along the same routes) based on the modelled impact of delay an operator causes to itself. Payments between operators are facilitated through the 'star model'⁶⁵ with Network Rail at the centre. This effectively means that Network Rail acts as a 'clearing house' for the financial payments associated with delays caused by one operator affecting others.⁶⁶

The Schedule 8 regime is slightly different for other types of train operators. Freight operator performance is defined in minutes of delay per 100 miles, whilst charter operator performance is based on average past performance over the 5-year recalibration period. For both freight and charter operators, payment rates are based on the average estimated financial impact of an operator causing a minute of delay to another train operator (rather than on the impact of delay caused to itself modelled on other operators, as in the passenger regime). The payment rate does not vary between operators in the freight and charter regimes.

Delay Repay

Schedule 8 does not include any direct payments made to customers for train delays. The rail sector has its own delay compensation arrangements for this purpose, e.g. through the Delay Repay scheme which is divided into two categories. DR15 provides compensation for passengers whose train was delayed between

⁶³ Office of Rail and Road (April 2019): "Performance regime", p. 3-5 [[online](#)].

⁶⁴ Office of Rail and Road (April 2019): "Performance regime", p. 4 [[online](#)].

⁶⁵ The 'star model' effectively refers to the 'clearing house' mechanism for the financial payments associated with delays as initially all payment are made to Network Rail who then distributes these to the relevant operators. Further details are available [[online](#)].

⁶⁶ Network Rail: "Technical overview: Payments relating to disruption" [[online](#)].

15-29 minutes, while DR30 provides compensation to those customers who experienced delays of more than 30 minutes.⁶⁷

Electricity

Broad Measure of Customer Satisfaction

In DPCR5 Ofgem introduced a customer service incentive called Broad Measure of Customer Satisfaction (BMCS). Its aim is to incentivise network companies to deliver good customer service by replicating the types of measures which consumer-facing businesses usually face in a competitive environment. It consists of the following three components:

- Customer Satisfaction Survey;
- Complaints Metric; and
- Stakeholder Engagement Incentive.

The **customer satisfaction survey** aims at incentivising Distribution Network Operators (DNOs) to improve customer service by rewarding those who perform well while penalising those who do not.⁶⁸ Customer satisfaction is captured for three categories:

- Connections: aimed at customers who have received either a connection quotation or had a completed connection.
- Interruptions: aimed at customers who experience supply interruptions, either planned or unplanned.
- General enquiries: aimed at customers who have raised a general enquiry to their DNO.

For each of the above categories, the survey will ask customers to evaluate in a scale of 1-10, with 10 being the best mark, how satisfactory was the service they received from their DNO. The scores will then be averaged and a penalty or reward will be issued depending how well the DNO did compared to the target score. The payments are capped at +/- 1 per cent of the DNO's annual based revenue.

Similarly, the **complaints metric** aims at encouraging DNOs to manage customer complaints efficiently and to resolve these in a satisfactory manner.⁶⁹ In cases where DNOs do not meet the set targets for this measure, they could be penalised for up to 0.5 per cent of base revenue. There are four indicators against which complaints performance is measured with different weights attached to each indicator:

- Complaints unresolved after one day (10 per cent);
- Complaints unresolved in 31 days (30 per cent);
- Repeat complaints (50 per cent); and
- Number of Energy Ombudsman decisions that go against the DNO (as a percentage of total complaints) (10 per cent).

Performance on each indicator is weighted based on the weights in parenthesis to derive the overall score, with a lower mark meaning a better score. Under the current price review, the target is 8.33 and the

⁶⁷ Office of Rail and Road (June 2020): "Consultation on improving access to delay compensation: Proposals for a Licence Condition, Delay Compensation Code of Practice, and Third Party Intermediaries Code of Conduct", p.5, [\[online\]](#).

⁶⁸ Ofgem (July 2020) "RIIO-ED2 Sector Methodology Consultation: Annex 1 - Delivering value for money services for consumers", p.16 [\[online\]](#).

⁶⁹ Ofgem (July 2020) "RIIO-ED2 Sector Methodology Consultation: Annex 1 - Delivering value for money services for consumers", p.28 [\[online\]](#).

maximum penalty score is 14.84. Until now, all DNOs have performed better than the target and thus no penalty has been incurred.

Finally, the **Stakeholder Engagement and Consumer Vulnerability incentive** aims at incentivising DNOs to engage with consumers and other relevant stakeholders in order to anticipate future needs.⁷⁰ It financially rewards companies which undertake high-quality engagement activities to inform their future business decisions. It is a reward-only incentive of up to 0.5 per cent of annual allowed revenues and it is determined using a panel of independent experts which will assist Ofgem in giving a score to each company. However, under the current review for the next price control, Ofgem found that high-quality engagement, while still highly relevant and important, is now a business-as-usual activity for DNOs and as such no incentive mechanisms is required to encourage high-quality customer engagement. Therefore, they suggest removing this incentive.

Quality of service incentives

The **Interruptions Incentive Scheme (IIS)** under the current price control aims at incentivising DNOs to improve their networks' overall reliability by setting target levels. It covers all interruptions which are of three minutes or longer irrespective of whether these are planned or unplanned.⁷¹ Part of this scheme is the **quality of service incentive**, a financial incentive which measured DNOs performance against:

- The number of customers interrupted per 100 customers (CI)
- The number of customer minutes lost (CML).⁷²

Under CI, 1.2 per cent of revenue is exposed while this figure is 1.8 per cent under CML. The amount of revenue exposure has been informed by a customer survey.

Another standard in the electricity market is the **Guaranteed Standards of Performance (GSoPs)**.⁷³ GSoPs set the minimum levels of performance which DNOs have to achieve and in the case where they do not meet the service levels, payments will be made to the affected customers. Under the GSoPs, in the cases where DNOs do not meet their target, they should compensate the affected customers even if the latter do not make a claim. The level of payments and the standards are reviewed ahead of every price control and are adjusted when necessary. In addition, the payments under the GSoPs standard can differ between domestic and non-domestic customers.⁷⁴

Worst Served Customers (WSC) mechanism is another incentive aimed at addressing the experience of customers who might have not been (fully) covered by the IIS, especially if they faced an unusually high number of interruptions.⁷⁵ The mechanism operates as a use-it-or-lose-it allowance incentivising to improve reliability for those customers who experience very poor levels of service.

⁷⁰ Ofgem (July 2020) "RIIO-ED2 Sector Methodology Consultation: Annex I - Delivering value for money services for consumers", p.33 [\[online\]](#).

⁷¹ Ofgem (July 2020) "RIIO-ED2 Sector Methodology Consultation: Annex I - Delivering value for money services for consumers", p.68 [\[online\]](#).

⁷² Ofgem (n.a.) "Quality of Service Incentives" [\[online\]](#).

⁷³ Ofgem (July 2020) "RIIO-ED2 Sector Methodology Consultation: Annex I - Delivering value for money services for consumers", p.97-98 [\[online\]](#).

⁷⁴ See 'Non-connections Guaranteed Standards of Performance' on p.163-164 in Ofgem (July 2020) "RIIO-ED2 Sector Methodology Consultation: Annex I - Delivering value for money services for consumers" [\[online\]](#).

⁷⁵ Ofgem (July 2020) "RIIO-ED2 Sector Methodology Consultation: Annex I - Delivering value for money services for consumers", p.101-102 [\[online\]](#).

Aviation

Airport regulation also includes service quality regulation in a setting where end users (passengers) will experience a service that depends on the actions of many different providers, including the airport. Economic regulation is focussed on the airport. CAA has included a Service Quality Rebates and Bonuses (SQRB) scheme since 2003. This currently covers five broad areas at Heathrow airport: passenger satisfaction with airport cleanliness, departure lounge seating, information and directions; security queue management; availability of passenger-facing equipment, such as lifts; availability of airline-facing equipment, such as stands; and aerodrome congestion.

For the forthcoming Heathrow price cap (Q7), the CAA has decided to update its approach to realise an approach more focussed on outcomes rather than outputs.⁷⁶ The CAA cites the approach in the water sector as a lesson from which it hopes to learn in developing a more outcomes focused regime.

In thinking about the design of an incentive regime, the CAA observes that one risk with outcomes-based approaches is that the airport may not be the only party responsible for the passenger experiencing a bad outcome. But the CAA also notes that such scenarios do not automatically imply that the airport should be freed from consequences. It cites an example of Gatwick airport reporting on the performance of airlines and ground handlers as a way that an airport might incentivise better effort by other parties at the airport so as to realise good outcomes for passengers. In a competitive setting, companies would risk losing market share if they made no effort to improve the behaviour of other suppliers whose actions affect the outcomes their passengers experience. Airports whose passengers always have to wait a long time to receive their baggage on arrival at an airport risk losing their custom to rival airports if they face meaningful airport competition, whether it is employees of the airport, an airline or a ground handler who is actually to blame for the delay.

The CAA's design principles for outcome-based service quality regulation have five principles.⁷⁷

- Robust consumer research should inform the OBR, with this work led by the airport. This is consistent with the approach favoured by Ofwat for PR14 which CAA alludes to in its policy paper.
- 'Outcomes', 'measures', 'targets', and 'incentives' should all feature in OBRs
- The airlines and the Consumer Challenge Board (CCB) should play an important role in developing OBRs
- It should build on the pre-existing SQRB scheme
- The target for performance reporting should be consumers

From a design perspective, the second of these principles is perhaps the most relevant. **Outcomes** should summarise the most important aspect of the airport services that consumers value, expressed in an easy way for consumers to understand. They can coincide with the interests of other stakeholders but they must be important to consumers. For each outcome, there should be one or more performance **measure** to track progress against. The overall suite of measures should cover all aspects that are important, whether directly or indirectly, to realising the outcomes consumers care about. Even if a consumer is unaware about some aspect of airport operations, if it is fundamental to realising outcomes that passenger care about then it is an important aspect of service. **Targets** should be set having regard to customer preferences, the scope for management to improve performance and the willingness to pay of consumers and airlines to realise improvements above and beyond what is possible using existing facilities. **Incentives** should be justified and calibrated. Generally, they are expected to include a financial component (which may include both bonuses and penalties).

⁷⁶ CAA (2016): "Future of service quality regulation for Heathrow Airport Limited: Consultation on the design principles for a more outcome-based regime" [[online](#)].

⁷⁷ CAA (2016): "Future of service quality regulation for Heathrow Airport Limited: Consultation on the design principles for a more outcome-based regime" [[online](#)].

Heathrow Airport published an Initial Business Plan ahead of the forthcoming price cap. The feedback from the CCB and the CAA was that the consumer engagement was generally good, but that the plan was missing 'the golden thread' linking the feedback from consumers with what the airport was actually planning to do. The initial business plan was prepared before the consequences of Covid-19 were known. More recently, HAL has published a revised business plan that CAA will be reviewing in the coming months.

The arrangements in airport regulation do not involve anything akin to the split between household and non-household customers. The SQRB scheme and the proposed OBR both would govern the entirety of the airport's regulated airport services.