IFF Research: BR-MeX Final Presentation

Contents

IFF Research

Introduction

IFF Research

Background & context

R-MeX: Retailer Measure of Experience

R-MeX is a survey that provides qualitative and quantitative measurement of how water retailers perceive the service of the wholesalers that they work with and has run twice a year since August 2018.

B-MeX: Business Measure of Experience

Ofwat's monitoring of the business retail market and its review of incumbent support for effective markets found that wholesalers need to improve their support for the non-domestic market. In particular, it was found wholesalers need a stronger focus on understanding the needs and requirements of their business customers and to tailor their wholesale service offerings accordingly.

Off the back of this, there have been calls from the water industry for there to be a Business Customer Measure of Experience (B-MeX), to measure the quality of services delivered by wholesalers to business customers in the business retail market. Work had already been conducted into the concept and feasibility of a B-MeX incentive, and a working group was set up to develop a model for a B-MeX survey.

BR-MeX: Business Retailer Measure of Experience:

Ofwat and MOSL intend to use the findings of both surveys to incentivise wholesalers to provide a good level of service to their business customers and water retailers.

IFF Research

BR-MeX Pilot: Objectives

Overarching aim: design, test and pilot all aspects of the B-MeX and R-MeX surveys to ensure that the survey methodology is robust for a shadow run in 2024, and full implementation from April 2025

The specific objectives of the research were to:

1. Design a pilot methodology for B-MeX and R-MeX surveys.

2. Conduct fieldwork to pilot and test both surveys and reflect on findings.

3. Analyse findings and produce recommendations and guidance for a shadow year of B-MeX and R-MeX fieldwork in 2024.

These activities will provide a base to inform Ofwat's decisions for both the BR-MeX Shadow period (2024-25) and final incentive design for when BR-MeX becomes operational as from April 2025.

www.iffresearch.com

OFFICIAL

IFF Research

Approach summary

JULY-AUG	AUG-OCT	OCT-DEC	DEC-JAN	FEB-MAR
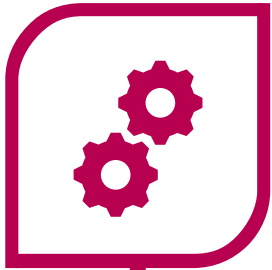
Scoping phase

To get a firmer understanding of key questions to inform our overall approach and the design of the pilot surveys

Cognitive testing

50 with business customers to test the B-MeX survey

10 with retailers to test the R-MeX survey.

Tested with a range of participant types

Pilot exercise

1,408 CATI interviews were conducted to test the B-MeX survey.

17 of 20 retailers completed the R-MeX survey.

Follow-up qual

25 follow-up qualitative depth interviews.

To check how respondents found the questions and explore any additional issues that arose during the pilot.

Analysis and reporting

Presentation of key results and recommendations.

Final report to be published online (March/April 2024)

Guidance documents for shadow runs of surveys in 2024-25.

www.iffresearch.com

OFFICIAL

IFF Research

Business Measure of Experience

B-MeX Pilot Analysis

IFF Research

Sample quality and outcomes

A total of 18,563 records were received from water wholesalers and MOSL for the B-MeX pilot

Sample from water wholesalers is known as 'direct sample.' Sample from MOSL is known as 'indirect sample,' and relates to business customer requests that retailers actioned – i.e. requested a wholesaler service on behalf of the customer – via the Bilateral Hub.

The sample included direct and indirect customer contacts that occurred over the period 18.09.23 to 18.10.23

	Direct Sample	Indirect sample	Total sample	% of overall useable sample
Affinity Water	258	58	316	2%
Anglian Water	1,633	296	1,929	10%
Northumbrian Water	794	123	917	5%
Portsmouth Water	83	9	92	1%
SES Water	90	21	111	1%
Severn Trent Water	6,252	788	7,040	38%
South East Water	548	129	677	4%
South Staffordshire Water	144	83	227	1%
South West Water	924	245	1,157	6%
Southern Water	826	96	922	5%
Thames Water	905	422	1,327	7%
United Utilities	958	325	1,283	7%
Wessex Water	898	40	928	5%
Yorkshire Water	1,512	113	1,625	9%
Total	**15,825**	**2,738**	**18,563**	**100%**

South West Water and Bristol Water are reported together for the purposes of the sample analysis.

www.iffresearch.com
OFFICIAL

IFF Research

Almost half of the records received were excluded

	% of total wholesaler sample excluded
Affinity Water	31%
Anglian Water	76%
Northumbrian Water	51%
Portsmouth Water	61%
SES Water	57%
Severn Trent Water	51%
South East Water	16%
South Staffordshire Water	81%
South West Water	46%
Southern Water	21%
Thames Water	56%
United Utilities	33%
Wessex Water	42%
Yorkshire Water	38%
Total	**48.14%**

After sample was received, IFF undertook a manual process in which all contact data was collated and cleaned to remove unusable sample. Records excluded consisted of:

No telephone number (4,692)

Duplicates (i.e. where a business had contacted their wholesaler more than once within the eligible fieldwork period) (2,488)

Date was not in scope of pilot fieldwork (1,540)

Exclusions under Section 1.3 of Ofwat's Request for Information for Pilot Data (216)

Wholesalers that were not in scope of the research (1)

IFF Research

A starting pilot sample of 9,627 records remained

'Useable' records consisted of any records that were not excluded after sample cleaning and these remaining records therefore formed the pilot starting sample.

The split by wholesaler, and direct / indirect, are shown in the table on the right.

	Direct Sample	Indirect sample	Total sample	% of overall sample	% of business customers in wholesaler region
Affinity Water	187	30	217	2%	5%
Anglian Water	407	64	471	5%	9%
Northumbrian Water	381	66	447	5%	8%
Portsmouth Water	30	6	36	0%	1%
SES Water	31	17	48	1%	1%
Severn Trent Water	3,266	196	3,462	36%	15%
South East Water	528	39	567	6%	4%
South Staffordshire Water	2	41	43	0%	3%
South West Water	510	118	628	7%	6%
Southern Water	675	50	725	8%	4%
Thames Water	460	130	590	6%	15%
United Utilities	652	203	855	9%	13%
Wessex Water	535	3	538	6%	3%
Yorkshire Water	945	55	1,000	10%	10%
Total	**8,609**	**1,018**	**9,627**	**100%**	**100%**

IFF Research

Nearly two fifths of records in the starting sample were marked poor quality

	Total poor records	% of wholesaler sample
Affinity Water	98	45%
Anglian Water	-	-
Northumbrian Water	43	10%
Portsmouth Water	2	6%
SES Water	-	-
Severn Trent Water	2,731	79%
South East Water	257	45%
South Staffordshire Water	-	-
South West Water	176	28%
Southern Water	7	1%
Thames Water	328	56%
United Utilities	-	-
Wessex Water	3	1%
Yorkshire Water	-	-
Total	**3,645**	**38%**

During the sampling process, **some records were flagged as 'poor quality.'** These were records that had an unclear reason for contact, **and** either no named contact or company name on the sample.

'Poor' records were not removed for the purposes of the pilot to help understand the impact of sample quality on response rates. **No indirect sample records were marked as poor** and therefore all 'poor' records came from the direct sample.

Overall, **38% of usable sample records were flagged as being 'poor,'** although it should be noted that the majority (75%) of these records were from Severn Trent. If Severn Trent's sample were excluded entirely, 'poor' records would constitute 14.83% of the usable sample.

IFF Research

Overall, 15% of dialed records completed the survey

There were several points at which respondents could leave the survey before completing:

- **Screen out due to being ineligible** for the survey (e.g. not a business orgainsation)

- **Refusal** to take part in the survey at the start ("Refused")

- **Drop out** having initially agreed to take part ("Partial interview")

	Complete	Partial	Refused	Ineligible	Unobtainable	No definite outcome
Affinity Water	23%	1%	29%	7%	4%	37%
Anglian Water	23%	1%	22%	4%	9%	41%
Northumbrian Water	17%	2%	23%	6%	6%	47%
Portsmouth Water	31%	-	14%	3%	6%	47%
SES Water	38%	-	13%	9%	-	40%
Severn Trent Water	7%	3%	11%	13%	4%	62%
South East Water	7%	2%	21%	4%	27%	39%
South Staffordshire Water	23%	2%	14%	2%	2%	56%
South West Water	11%	2%	17%	3%	18%	48%
Southern Water	21%	2%	21%	8%	6%	43%
Thames Water	20%	2%	20%	7%	7%	45%
United Utilities	21%	1%	16%	5%	5%	51%
Wessex Water	23%	2%	19%	7%	5%	44%
Yorkshire Water	24%	2%	16%	6%	6%	46%
Total	**15%**	**2%**	**16%**	**9%**	**7%**	**51%**

www.iffresearch.com
OFFICIAL

IFF Research

Sample records classified as 'poor' were much less likely to complete the survey.

	% of 'poor' records dialled	% of 'Better' records dialled
Completes	7%	19%
No definite outcome	58%	47%
Refused	13%	18%
Ineligible	12%	6%
Unobtainable	6%	8%
Partial interviews	2%	2%

'Poor' records were also more likely to be classified as ineligible compared with non-poor records (12% vs 6%), however poor records were less likely to refuse to take part than non-poor records (13% vs 18%).

In addition, 'Poor' records were also more likely to be unreachable resulting in the survey not being completed (53% poor records vs 41% 'non-poor' records).

IFF Research

Outcomes: refusals

Overall, 16% of sample records dialed refused to complete the survey. The table on the right shows a breakdown of these reasons for refusal.

A high proportion (30%) were unable to recall any contact with their wholesaler.

Relatively few respondents refused to take part on the basis of not being able to speak in enough detail about the contact (8%), or not being clear on the distinction between wholesaler and retailer (1%).

	Number of refusals	% of refusals
No recollection of contact (S1/S2)	460	30%
Unspecified reason (S1/S2)	345	22%
Third party reported the issue on their behalf (S1/S2)	145	9%
Can't speak in enough detail about contact (S1/S2)	130	8%
Does not want to take part in any survey (S1/S2)	90	6%
They are a contractor who reported the issue (S1/S2)	66	4%
No longer works at the business (S1/S2)	66	4%
Company policy (S1/S2)	63	4%
Not happy to proceed with survey (A1)	46	3%
Not happy to proceed based on having to talk about interactions with wholesaler (S5)	25	2%
Not happy to carry on after definitions of wholesaler and retailer (S4)	12	1%
Taken part in recent survey (S1/S2)	10	1%
Another reason for refusal (S1/S2)	82	5%
Total	**1,540**	**100%**

www.iffresearch.com

OFFICIAL

IFF Research

A total of 1,408 completes were achieved for the pilot

Our statistician calculated ideal targets per wholesaler to allow robust analysis (were the survey run over 12 months) based on the sample provided; however, no hard quotas were set to allow completes to naturally fall out and test the performance of the sample provided.

Some wholesalers provided large volumes of sample, primarily due to their size, so completes by wholesaler were monitored to ensure we had a sufficient base for analysis amongst every wholesaler, particularly smaller wholesalers with fewer sample records.

	Ideal target	Completes	% of overall completes	% of business customers in wholesaler region
Affinity Water	34	49	3%	5%
Anglian Water	218	104	7%	9%
Northumbrian Water	102	72	5%	8%
Portsmouth Water	30	11	1%	1%
SES Water	34	18	1%	1%
Severn Trent Water	145	241	17%	15%
South East Water	30	41	3%	4%
South Staffordshire Water	79	10	1%	3%
South West Water	25	70	5%	6%
Southern Water	111	146	10%	4%
Thames Water	109	116	8%	15%
United Utilities	122	178	13%	13%
Wessex Water	134	124	9%	3%
Yorkshire Water	105	228	16%	10%
Total	**1,400**	**1,408**	**100%**	**100%**

www.iffresearch.com

OFFICIAL

IFF Research

Lower quality records achieved a lower response rate

A higher response rate was achieved with 'better' quality records (19%) compared with 'poor' quality sample records (7%), suggesting that **sample quality is critical for achieving good response rates.**

	'Poor' records response rate	'Better' quality records response rate
Affinity Water	24%	22%
Northumbrian Water	7%	18%
Severn Trent Water	6%	11%
South East Water	4%	10%
South West Water	8%	12%
Thames Water	18%	22%
Total	**7%**	**19%**

Note: Some wholesalers are not included due to low base size for comparison or for having no 'poor' records.

	Total sample dialled	Completes	Response rate
Direct	8,504	1,188	14%
Indirect	995	220	22%
Total	**9,499**	**1,408**	**15%**

Indirect records, provided by MOSL, achieved a higher response rate (22%) than direct sample (14%).

This is likely due to the indirect sample being higher quality with all data fields fully populated

IFF Research

Pilot survey findings & qualitative feedback

Most respondents in the qualitative interviews said they felt confident in the distinction between a water wholesaler and a water retailer

Some participants reported initially feeling unclear, but after their contact, or after having been read the definitions in the survey, felt the distinction was clear:

> *"Before my dealings with the wholesaler I was unclear. Afterwards I was much clearer and felt happy with the distinction."*
>
> ***Direct contact, Affinity Water***

> *"I just wanted to be clear on who was responsible for what... once it was explained it was perfectly understandable."*
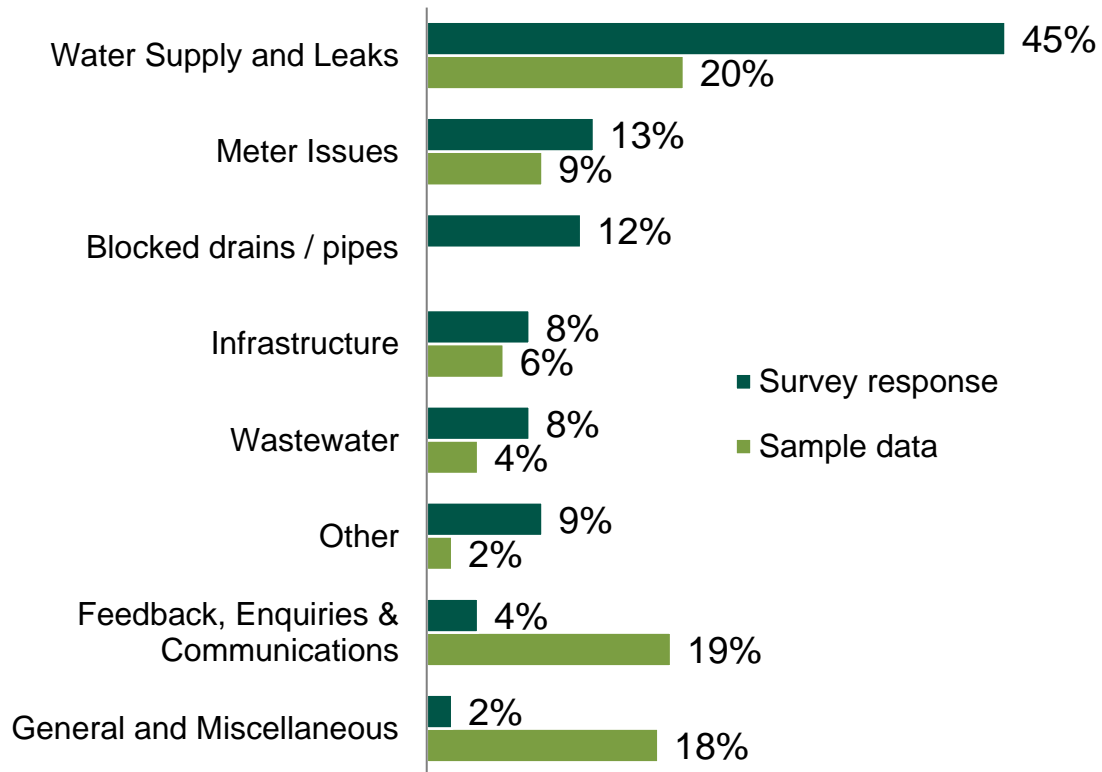>
> ***Direct contact, United Utilities***

Although many felt the definitions were clear, interviewers in the quantitative survey did report that sometimes they **spent considerable time before beginning the survey ensuring that the respondent was referring to contact with their wholesaler rather than their retailer.**

Looking forward to future iterations of the survey, making **the name of the respondent's retailer and wholesaler available to interviewers** is something to consider, to provide additional context on the interaction and to **avoid confusion.**
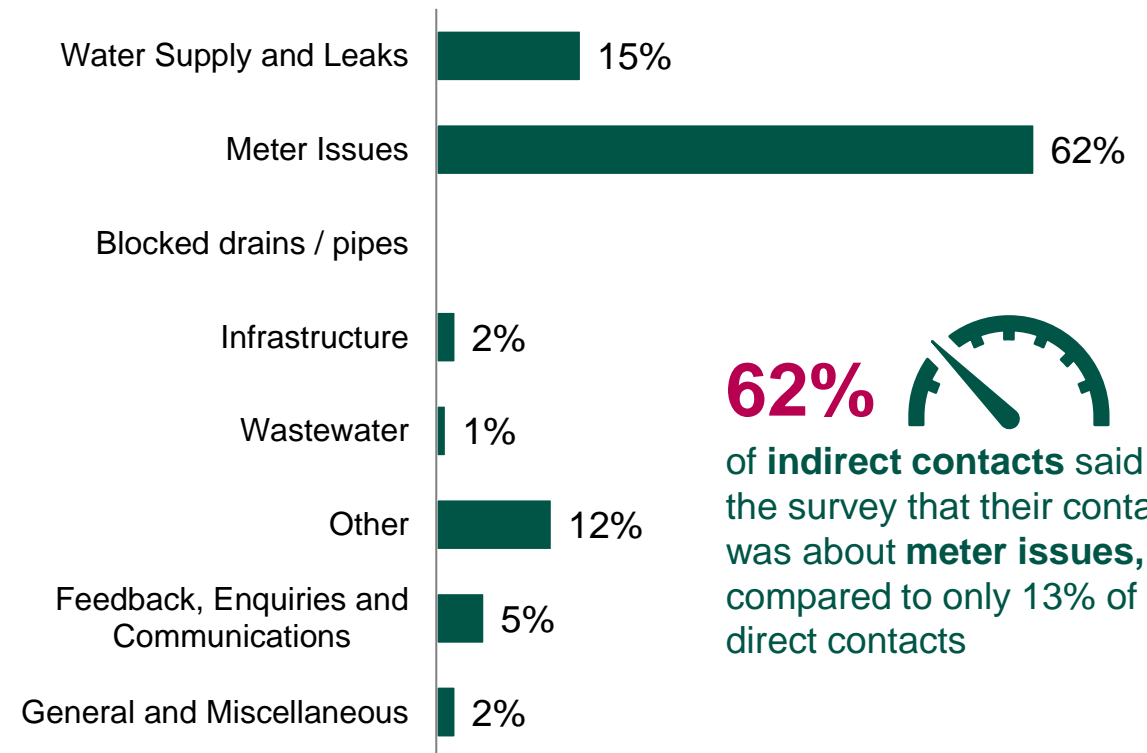
www.iffresearch.com

IFF Research

The reasons for contact given by direct contacts in the survey often differed from the reasons for contact listed on the sample file.

Reason for contact (direct contacts – sample vs. survey)

Reason	Survey response	Sample data
Water Supply and Leaks	45%	20%
Meter Issues	13%	9%
Blocked drains / pipes	12%	
Infrastructure	8%	6%
Wastewater	8%	4%
Other	9%	2%
Feedback, Enquiries & Communications	4%	19%
General and Miscellaneous	2%	18%

- Survey response
- Sample data

Reason for contact (indirect contacts – survey)

Reason	Survey
Water Supply and Leaks	15%
Meter Issues	62%
Blocked drains / pipes	
Infrastructure	2%
Wastewater	1%
Other	12%
Feedback, Enquiries and Communications	5%
General and Miscellaneous	2%

62% of **indirect contacts** said in the survey that their contact was about **meter issues,** compared to only 13% of direct contacts

S8 What was the main reason for making contact with [WHOLESALER] on that occasion? Direct contacts: 1,188; Indirect contacts: 220

IFF Research

The majority reported contacting their wholesaler directly

- In the survey, **three-quarters of respondents (75%) said they contacted their wholesaler directly,** as would be expected given the majority of sample were direct contacts.

- **13%** contacted their retailer, who in turn asked their wholesaler to get in touch

- **6%** were contacted by their wholesaler directly

- **2%** contacted their retailer, who then passed on their wholesalers' details.

76%

three-quarters of the contacts experienced by respondents in the survey were by **telephone**

A3 How did the contact with [WHOLESALER] on [SERVICE DATE] come about? Base 1,408. A4 Did you get in touch with wholesaler by..? Base: 1,356.
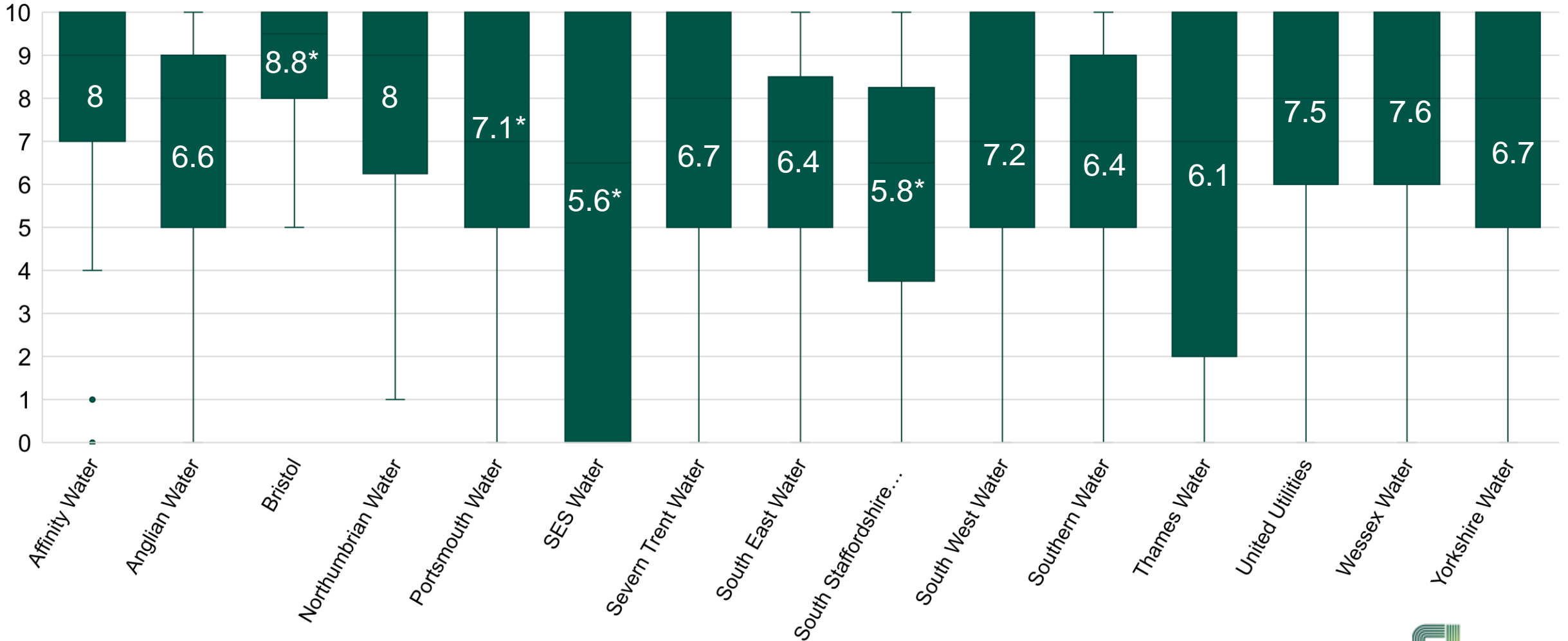
IFF Research

Satisfaction scoring in B-MeX

- Business customers were asked to give a satisfaction score for their contact with their wholesaler, using a 0-10 satisfaction scale

- Definitions were read out for the low, high and mid points of the scale to encourage consistency in scoring

"Thinking just about your contact with <Water_Wholesaler> on <Service_Date>, how satisfied are you with your experience of your contact?

Please use a scale of 0-10, where 0 = extremely dissatisfied, 5 = neither satisfied nor dissatisfied and 10 = extremely satisfied."

IFF Research

Two-thirds of respondents were satisfied (a score of 6-10 out of 10) with the contact they had with their wholesaler, while a fifth were dissatisfied (a score of 0-4 out of 10). The mean score across all respondents was 6.9.

Category	Score
Affinity Water	8
Anglian Water	6.6
Bristol	8.8*
Northumbrian Water	8
Portsmouth Water	7.1*
SES Water	5.6*
Severn Trent Water	6.7
South East Water	6.4
South Staffordshire…	5.8*
South West Water	7.2
Southern Water	6.4
Thames Water	6.1
United Utilities	7.5
Wessex Water	7.6
Yorkshire Water	6.7

IFF Research

In the qualitative interviews, there were mixed feelings on how easy it was to give feedback on the specific contact they had with their wholesaler

Some felt it was easy to talk about the contact they had with their wholesaler, and had no issues referring to the specific contact in question

"I found it really easy talking about it. I felt the survey gave enough chance to say what you wanted to say."

Direct contact, Yorkshire Water

One participant initially remembered it being relatively **straightforward**, but went on to mention that they had had multiple phone calls over a few days with their wholesaler, and that in reality it was **difficult to pinpoint the specific interaction**

Another felt that with **the limited information** the interviewer was able to give them, it was **extremely difficult** to figure out which contact they were talking about

"Unless I go back through my records, I can't determine what the contact was about. I'm making an assumption about what that contact was about, but it could be about this other matter."
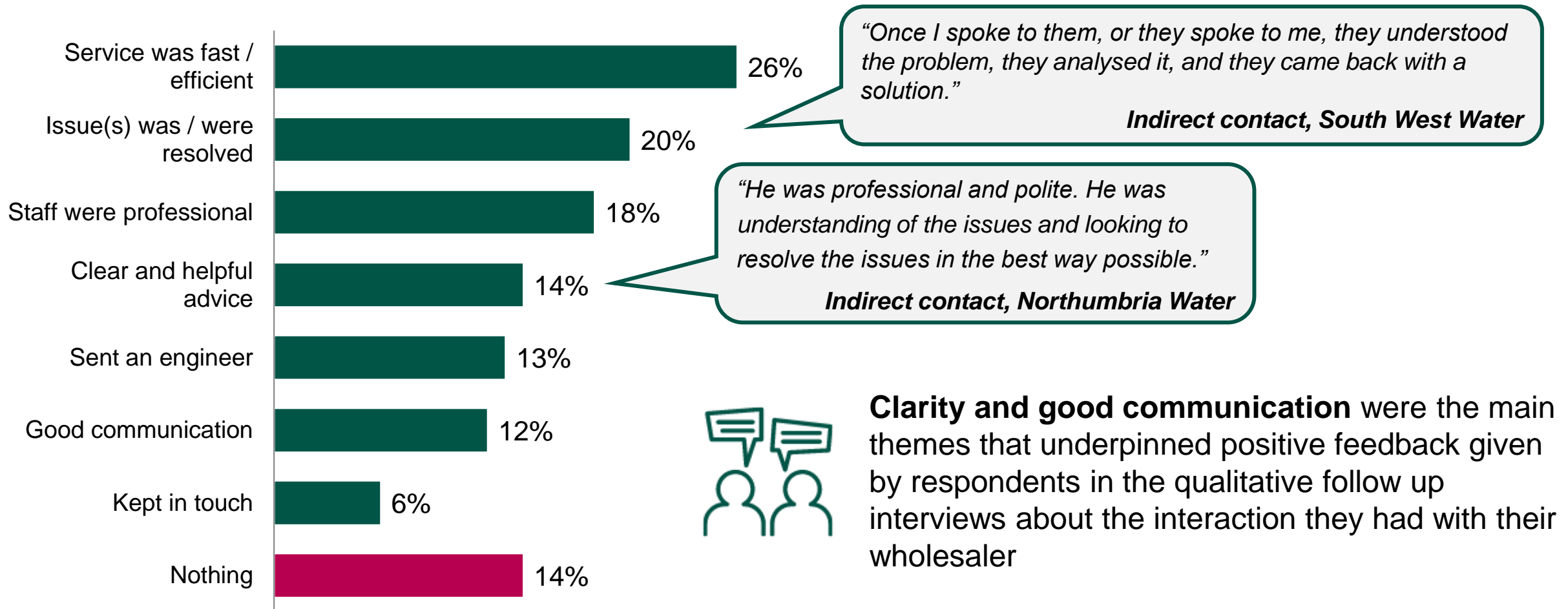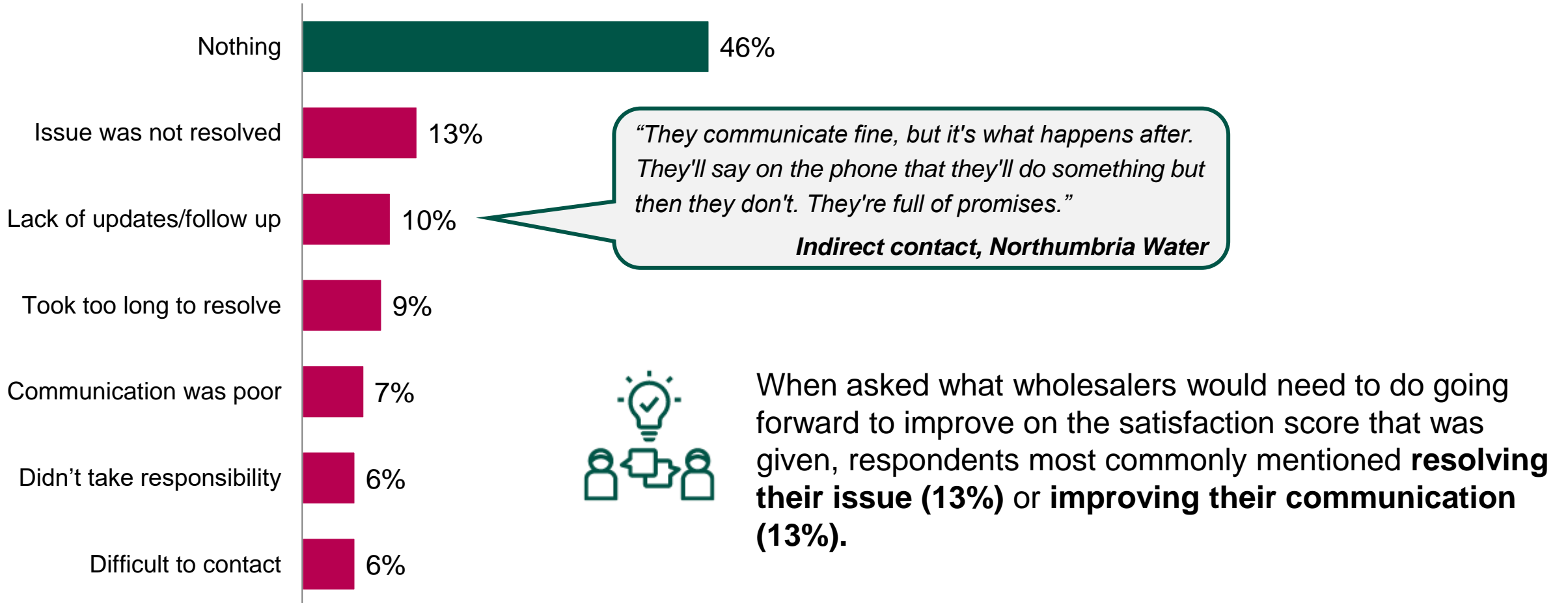
Indirect contact, South West Water

www.iffresearch.com

IFF Research

A fast and efficient service was appreciated by respondents, along with their issue being resolved

Category	Percentage
Service was fast / efficient	26%
Issue(s) was / were resolved	20%
Staff were professional	18%
Clear and helpful advice	14%
Sent an engineer	13%
Good communication	12%
Kept in touch	6%
Nothing	14%

> *"Once I spoke to them, or they spoke to me, they understood the problem, they analysed it, and they came back with a solution."*
>
> **Indirect contact, South West Water**

> *"He was professional and polite. He was understanding of the issues and looking to resolve the issues in the best way possible."*
>
> **Indirect contact, Northumbria Water**

Clarity and good communication were the main themes that underpinned positive feedback given by respondents in the qualitative follow up interviews about the interaction they had with their wholesaler

www.iffresearch.com

A8 Just thinking about your contact with [WHOLESALER] on [SERVICE DATE] what would you say they did well? Base: 1408

IFF Research

Nearly half of respondents said that there was nothing their wholesaler didn't do well during their contact. Some felt there was a lack of follow up after their contact, and others were disappointed that their issue was not resolved

Category	Percentage
Nothing	46%
Issue was not resolved	13%
Lack of updates/follow up	10%
Took too long to resolve	9%
Communication was poor	7%
Didn't take responsibility	6%
Difficult to contact	6%

> "They communicate fine, but it's what happens after. They'll say on the phone that they'll do something but then they don't. They're full of promises."
>
> **Indirect contact, Northumbria Water**

When asked what wholesalers would need to do going forward to improve on the satisfaction score that was given, respondents most commonly mentioned **resolving their issue (13%)** or **improving their communication (13%).**

IFF Research

B-MeX
Recommendations

IFF Research

The survey should be conducted over the phone

Feedback

Some customers are not likely to confidently understand the difference between their retailer and wholesaler without the definition read out by the telephone interviewer.

This definition could be provided as text in an online survey however participants will likely skip over larger blocks of text. Telephone interviewers are also able to take a live 'quality control' role, by not proceeding with interviews if it is clear that the respondent does not understand the role of their wholesaler specifically.

Recommendation

We recommend that B-MeX is conducted as a telephone survey. There are two key reasons for this:

- A telephone methodology will provide **more accurate and relevant responses** – this will be critical for the mainstage of the survey when financial incentives and penalties be based on the scores.

- Telephone surveys also **achieve better response rates** than online surveys, which will ensure the results are **more robust**, including for smaller wholesalers who will have less sample available to survey.

Feedback

Telephone interviewers also played a valuable role helping prompt respondents to remember the specific contact which we were seeking their feedback on. There is a risk that with an online survey respondents might skip this information and give feedback about a different contact or give more general feedback about their water supply.

www.iffresearch.com

OFFICIAL

IFF Research

Sample screening and quality needs to improve

Feedback

Analysis of the pilot shows that **sample quality has a clear impact on response rates** and the **ability of respondents to recall and give feedback** on specific contacts with their wholesaler.

Additionally, poor quality sample has further implications on:

- Costs;
- Fieldwork delays;
- Processing time; and
- Quality, quotas and weighting.

Recommendation

Sample from all wholesalers should have ineligible contacts excluded (e.g. passers-by). Inclusion of ineligible contacts would lead to wasted interviewing time, damaging response rates, and make running the survey more expensive.

Recommendation

Sample files must contain all data fields deemed essential, and essential fields must be fully populated and in the correct format. Non-essential fields should also be populated, and in the correct format – these will help to improve response rates and ensure the results are representative.

www.iffresearch.com
OFFICIAL

IFF Research

Clear, understandable info on contact

Feedback

Some respondents found it **difficult to differentiate** between their wholesaler and retailer, and some found it **difficult to recall** the exact contact being referred to by the interviewer.

In cases where the sample contained relevant information (e.g. a clear reason for contact, and the name and business name the contact involved), interviewers were able to prompt respondents to aid their recall and understanding. However, where sample records were **missing this information**, or where information such as the 'reason for contact' used **internal jargon**, it was **difficult to ensure respondents were giving relevant feedback**.

Recommendation

Primary and Secondary (detailed) reason for contact codes/text should be phrased so customers and interviewers will understand. Internal codes and jargon should be avoided as much as possible, or used alongside a 'layperson' explanation.

Recommendation

Indirect sample should include a field with information on the interaction between the customer and wholesaler, equivalent to 'primary reason for contact.'

www.iffresearch.com
OFFICIAL

IFF Research

Quotas & weighting to ensure representative data

Feedback

The aim of the B-MeX survey is to derive a robust score for each wholesaler. To ensure the data is robust we will need to:

- Conduct **a sufficient number of interviews per wholesaler,** to ensure a good base for analysis

- Ensure the **completed interviews are representative of the population**

This can be achieved by **setting quota targets** during fieldwork, and **weighting the final data**. Weighting is a technique used to adjust the results of a survey to make it **more representative of the larger population**. The weights are used to balance out the results and provide a more accurate picture of the population's views.

Recommendation

Set quotas during fieldwork to ensure the achieved interviews are as close to the target profile as possible. Overall quotas should be set by wholesaler, and within wholesaler quotas could be set by direct/indirect contact, type of contact, outcome, and reason for contact (if harmonised data available)

Weight the data to population targets by wholesaler; again, these will require complete, harmonised sample data across wholesalers.

Recommendation

We would suggest **reviewing and revising the strategy for setting quotas and subsequent weighting during the shadow year**, when hopefully wholesalers will have been able to harmonise and improve the sample data supplied for B-MeX.

OFFICIAL

IFF Research

Harmonisation of sample

Feedback

In order to set targets for fieldwork and for weighting, wholesalers will need to deliver sample files with **complete information** about each contact – this will allow us to determine the characteristics of the population we need to represent.

It will be important for wholesalers to use **clear, consistent categories** to ensure data is treated consistently, and to enable accurate processing of sample data.

Recommendation

The **recording and coding of contacts to be harmonized and made consistent** across different wholesalers, including **'primary reason for contact.'** This will be critical for setting quota targets and weighting data to be representative.

Recommendation

Ensure that all essential data fields are populated for every sample record, e.g. contact name, tel no., service type, date of contact, outcome of contact, primary reason for contact, secondary reason for contact.

IFF Research

Questionnaire should remain largely the same

Feedback

There was limited feedback on further questionnaire changes during the cognitive stage and pilot stages.

However, feedback from our interviewing team and quality control checks when listening in to interviews reveals there could be confusion between wholesaler / retailer in some cases.

Recommendation

We would suggest **making the name of the respondent's retailer available to interviewers** along with the name of the wholesaler, to provide additional context on the interaction and to avoid confusion.

Aside from this we believe that the questionnaire worked well in the pilot, and we do not recommend making any further changes.

www.iffresearch.com

OFFICIAL

IFF Research
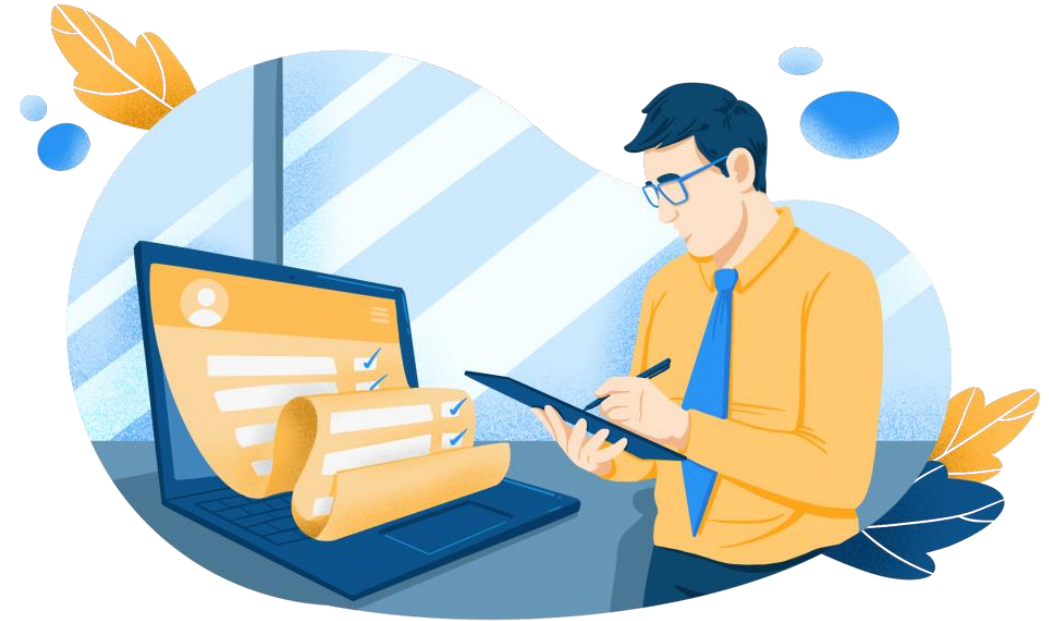
Retailer Measure of Experience

R-MeX Pilot Analysis

Cog testing high level findings: Retailers spend significant time and resource completing the survey

Scoring

Some retailers highlighted how they considered potential discrepancies in how different individuals in their organisation might score a wholesaler using the 0-10 scoring system.

These retailers took actions to prevent this by having **one individual responsible for moderating scores** submitted across their organisation to prevent inconsistent and anomalous scoring *within retailer*.

Responsibility

All retailers reported that the R-MeX survey is completed by multiple individuals within their organisation, but that **one individual has the ultimate responsibility** of leading on the completion and for collating responses.

Time

Retailers reported that completing the R-MeX survey is **time-consuming and resource intensive**.

However, this was caveated by most saying they are **willing to dedicate this time and resource** to the R-MeX survey because they understand why it is being conducted, why this is important, and **because they see the process as being of benefit to their own organisation**.

IFF Research

Changes implemented for pilot survey

01

Added labels to the existing 0-10 scoring system:
0 was labelled as 'very dissatisfied', 10 as 'very satisfied' and the mid-point as 'neither satisfied or dissatisfied'.

02

Redesigned the spreadsheet including the removal of macros to make the spreadsheet less restrictive to retailers when filling it out.

03

Added a catch-all optional free text response box (for the pilot survey only), to capture retailers' feedback on how they found the revised survey, and feed into post-pilot recommendations.

www.iffresearch.com

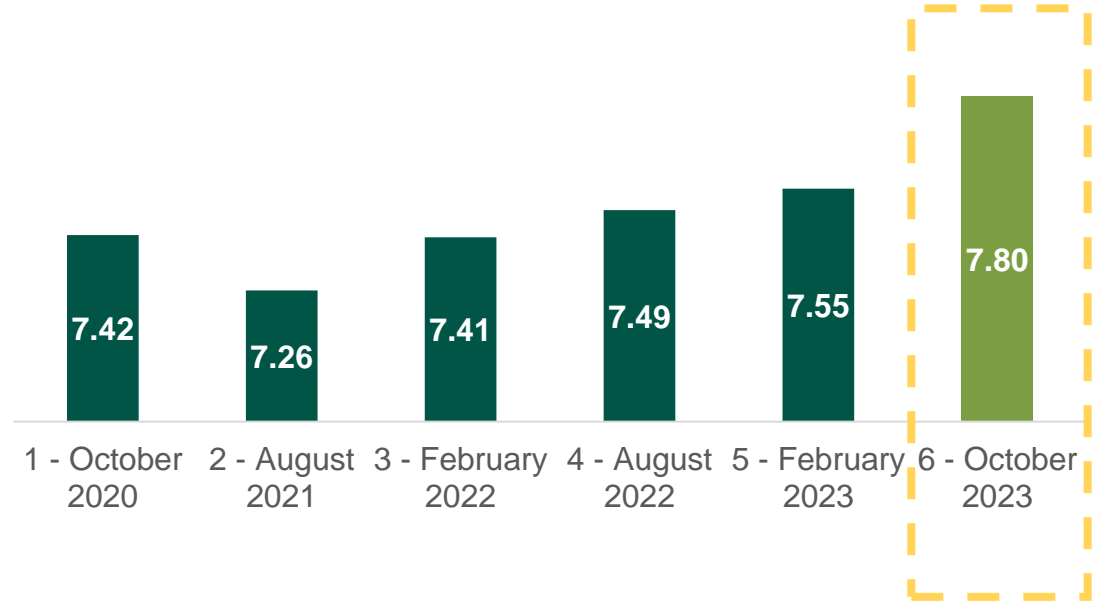IFF Research

Pilot survey results

Average scores for overall service and each of the 6 elements of service measured have broadly remained consistent or in line with upwards trends of previous waves since the launch of R-MeX.

The latest wave, which includes the changes made for the pilot on the scoring guidance, appears to mostly mirror this trend, and therefore suggests our revised approach including guidance is consistent with retailers' previous understanding for scoring wholesalers. However, there is still a possibility that retailers may be using different scoring styles.

Average score for overall service[1]

1 - October 2020	2 - August 2021	3 - February 2022	4 - August 2022	5 - February 2023	6 - October 2023
7.42	7.26	7.41	7.49	7.55	7.80

[1]This average score will be used by Ofwat & MOSL to help construct each wholesaler's performance score for BR-MeX

Base: October 2020 (n=14), August 2021 (n=14), February 2022 (n=13), August 2022 (n=16), February 2023 (n=15) and October 2023 (n=17)

www.iffresearch.com
OFFICIAL

IFF Research

Pilot survey results

However, there is concern that **if retailers do not all use the scale consistently** (e.g. if some retailers always give higher or lower scores than others, due to differences in interpretation of the scale or how levels of service should be ranked) **this could bias the scores for wholesalers** depending on which retailers they receive scores from

- Rating scales can, if not carefully moderated, **elicit different response styles** from different individuals
- These would tend to be **more noticeable in more granular scales**
- Over large samples, these differences 'even out' naturally
- However, as R-MeX only has a **small sample size**, such differences, if present, could have a bigger effect on wholesalers' final scores

For example, if Retailers A and B always tend to give higher scores, and Retailers C and D always tend to give lower scores, a wholesaler who is scored by only Retailers C and D could end up with a lower average score than a wholesaler scored by Retailers A and B (and a wholesaler scored by all four would be in the middle), even if their actual performance was the same.

www.iffresearch.com

OFFICIAL

IFF Research

R-MeX Recommendations

Maintain current method & frequency

Recommendations

- **Maintain current Excel methodology and survey frequency** – retailers felt both worked well in their current format.

- **Maintain current question wording as the vast majority of retailer feedback was positive -** questions were felt to be clear, easy to understand and easy to answer, and to cover relevant areas

- **Consider an alteration to the scoring approach for 'overall satisfaction' to ensure that any differing response styles across retailers do not unduly impact scores** – we present two options for consideration on the next slide, following a review of the data by a consultant statistician.

- *Note scores for individual elements of service will not feed into the final BR-MeX score and so do not need to be altered. Feedback to wholesalers will still include these scores, and qualitative feedback on what wholesalers are doing well or can do more to improve.*

www.iffresearch.com

OFFICIAL

IFF Research

Options for scoring system

Option 1: Adopt a less granular ~~ambiguous~~ scale with detailed descriptions

For example, four categories with a key to provide guidance on when to rate in each box.

SEGMENT	SCORE	Description
OUTSTANDING	1	TBD…
GOOD	2	TBD…
REQUIRES IMPROVEMENT	3	TBD…
UNSATISFACTORY	4	TBD…

For analysis we would suggest using a 'percentage in box' comparison.

As only the 'overall satisfaction' score will feed into the overall BR-MeX score, this would only need to apply for the overall score (scores for individual service aspects could remain consistent with previous waves).

Pros:
- forcing a limited number of options with more prescriptive instructions would facilitate consistent use of the measuring instrument and minimise any response style differences
- A more stable, robust method of comparison

Cons:
- A lack of discrimination between the performance of wholesalers due to the less granular scale

www.iffresearch.com

OFFICIAL

IFF Research

Options for scoring system

Option 2: Adjust the raw 'overall satisfaction' score to neutralise impact of any inconsistencies in retailer response style

Making this adjustment prior to averaging arguably enables a fairer comparison of average performance across wholesalers.

The suggested adjustment is to 'centre' the rating scores for each retailer (of each wholesaler). This is done by:
- For each retailer, calculating its average rating across scores for all wholesalers it rates
- Transform the score for each wholesaler by subtracting the mean from their raw score

Example:

- Retailer A has a **mean rating of exactly 7** (the average of the overall scores given to all wholesalers they rated)

- Retailer A scored Wholesaler X a score of 9.
  This is transformed to an **adjusted score of 2** (9 – 7 = 2)

- Retailer A scored Wholesaler Y a score of 6.
  This is transformed to an **adjusted score of -1** (6 – 7 = -1)

- The final score for each wholesaler is the **mean average of their adjusted score** for each retailer

Pros:
- Allows retention of the 0-10 rating scale
- Facilitates a fair comparison of the 'relative' ratings of wholesalers

Cons:
- The differences in 'absolute' score for a specific wholesaler given across retailers are lost.

IFF Research

Thank you
Any Questions?

04/03/2024

IFF Research