

Andrew Beaver  
Ofwat  
Centre City Tower  
7 Hill Street  
Birmingham  
B5 4UA

**Anglian Water  
Services Limited**

Anglian House  
Ambury Road  
Huntingdon  
PE29 3NZ

Tel: 01480 323047

Our ref:

30 January 2014

Dear Andrew

**Service incentive mechanism (SIM) for 2015 onwards**

You have our draft comments on the consultation sent in advance of the SIM Workshop held on Friday 24<sup>th</sup>, and our submission remains unchanged (attached for ease of reference).

Thank you for holding such an open and discursive workshop where thoughts and opinions were freely and openly considered.

Having taken the time to reflect upon the discussions and possible direction for SIM, we would like to share our strong view that there is either a real opportunity to build upon the success of SIM and make it a truly customer led measure, or a real risk that the measure becomes further complicated.

We urge you to consider two key issues:

**Qualitative only SIM -**

- A purely qualitative SIM as the only way to ensure that it is the customers' voice which determines incentive performance
- It avoids a heavy regulatory burden and inconsistencies in reporting
- It is easily understood by customers, media and other stakeholders alike, and can be used in cross industry comparisons
- It is truly outcome based
- It should evolve for AMP6, and we support the removal of notice, increase in frequency, and reduction in survey length. However, it should remain closed contacts only, as customers do not want to be surveyed until an issue is resolved

Registered Office  
Anglian House,  
Ambury Road, Huntingdon,  
Cambridgeshire. PE29 3NZ  
Registered in England  
No. 2366656

an AWG Company

and to do so reflects poorly upon the regulator and the company concerned.

**Quantitative -**

- The quantitative measure is 'old school' regulation, but if retained must not be expanded to include all contacts. To do so will stifle innovation – for example, we would seriously consider pulling back on all **live chat, social media** and **text messaging** activity. Areas in which we believe that the industry could already rightly be criticised for lagging behind other industries.
- The existing complexity of the quantitative measure would be multiplied and would therefore require significant guidance and clarification of definitions, followed by industry wide horizontal audits.
- The regulatory burden of managing such a measure correctly would be disproportionate and counter to the direction of lighter touch and outcome based regulation.
- The current method of calculation means that the intended 50:50 weighting between quantitative and qualitative is not realised. Any consideration of a revised weighting must also address bias in the calculation.

Gaining consensus on these issues is never easy, but we do encourage Ofwat to be bold, and make SIM a fully customer driven and focused measure. We have been in favour of a qualitative only SIM since the inception of the measure; indeed at a time when our own qualitative performance was lower quartile. Our position then, as it is now, is that a qualitative only SIM is the superior measure for the industry.

We remain strong supporter of SIM, so if we can assist further in any way, we would be happy to do so.

Yours sincerely

Jean Spencer  
Regulation Director

cc Price Review Programme Team

## **Appendix 1**

### **Overall design of the incentive**

- **Q1 Range for rewards and penalties – proposition that SIM should incentivise improvements in companies operating below the frontier and, whether further service improvement in frontier companies should be incentivised**

Companies performing in the upper quartile have over recent years pushed their performances to even higher standards. Incentives for being at the upper end of any performance league table are imperative to maintain this stretch. It challenges us to remain constantly vigilant and focused upon the service delivered.

It is important to recognise that with customer expectations continually rising, driven by experience in other sectors, maintaining a score within the league table requires service improvement in real terms.

When assessed as a percentage of the margin available to improve, those at the upper and lower ends of the qualitative league table have improved by very similar levels.

	<b>2010/11</b>	2011/12	<b>2012/13</b>	<b>Actual shift in score</b>	<b>Range available</b>	<b>% of range moved</b>
Averages						
1st-7th	<b>4.46</b>	4.54	<b>4.62</b>	0.16	0.54	<b>29.63%</b>
8th-14th	<b>4.28</b>	4.38	<b>4.48</b>	0.20	0.72	<b>27.78%</b>
15th-21st	<b>3.93</b>	4.11	<b>4.23</b>	0.30	1.07	<b>28.04%</b>

The consultation claims that SIM and OPA have been good at promoting catch-up improvements and poorer at promoting frontier improvements. Any performance improvement at the bottom end of the spectrum is likely to be easier than an equal improvement at the top due to diminishing marginal returns. The results above when seen in this light show remarkable improvements in the frontier performers' scores.

In the case of the OPA, we agree that frontier improvement was eventually less evident, but this was a function of assessing performance against defined performance levels which resulted in near perfect scores at the frontier. SIM does not suffer from this draw-back, and in fact was specifically introduced in order to provide an incentive which would drive the frontier forward.

Incentivising both ends of the spectrum does maintain focus and creates driven strategies to ensure service levels are matching, if not bettering, those outside the water sector.

**It is therefore our recommendation that incentives in their current form are maintained**

- **Q2 SIM should continue to incentivise service throughout the value chain, and not just those elements under the sole control of retail businesses.**

SIM has stimulated service delivery improvements and as such should remain a measure and incentive throughout the value chain and not solely focus on retail activity. How the individual companies manage this incentive throughout the business should be a matter for them to decide. How well the companies manage these arrangements will be evidenced through improving service deliverables.

**Our recommendation is that SIM should continue to apply throughout the value chain, and companies should be left to develop internal contracting arrangements as they see fit**

### **Detailed design questions**

As stated above, the evidence shows that the current mechanism with incentives at both ends of the league table is encouraging improvement throughout.

**Q3 Design issue 1 – symmetric or asymmetric incentive, retain current structure, or only downside of 1%, or symmetric of downside 1% and upside 1%**

The risk with the current asymmetric incentive is that innovation and progression at the upper range could slow. Any new innovation brings some risk and whilst this can be managed to mitigate adverse reaction the incentive to remain at the forefront must be retained. The introduction of a symmetric incentive between upside and downside would encourage companies to strive further, ensuring the industry becomes recognised as leading within the wider business community.

However, given the performance gains at the frontier which are still being made with the existing incentive, we see no compelling reason to change at this stage.

Cost-benefit analysis is not available for SIM improvements, and the existing structure was introduced without a detailed cost-benefit analysis. Marginally diminishing benefits would however suggest that an asymmetric incentive may be theoretically appropriate, but the degree of the asymmetry is unknown.

**Our recommendation is to retain the current asymmetric incentive – option 1**

- **Q4 Design issue 2 – maintain the current magnitude of financial incentive in the range of +0.5% to -1%**

The current range of incentive and financial exposure has been seen to drive improvement. Predicting customer trends and expectations is difficult and chasing every innovation could expose the company to unnecessary service risk. We believe the introduction of some degree of certainty at the top and bottom of the range would ensure service levels throughout the league table continue to improve in line with wider service industries and the customers changing expectations (e.g. the top performing company received the maximum reward, and the bottom received the maximum penalty).

The current magnitude of incentive has been a feature of the regulatory system for many years under SIM and OPA. We see no reason to change the magnitude at this stage, especially when companies have been given the opportunity to suggest other outcome delivery incentives in their business plan submissions.

**We support the proposal to broadly maintain the current magnitude of financial incentive.**

- **Q5 Design issue 3 – balance and weighting of the qualitative and quantitative measures**

We remain a strong supporter of the qualitative element of the mechanism as it reflects the customer's voice:

- it is a demanding measure of customer satisfaction
- requires consistency of service
- is easily understandable by customers, stakeholders and the media
- allows direct cross utility and sector comparison as it measures customer satisfaction.

We believe the qualitative measure to be the most meaningful of the customer experience, as it measures how customers rate water services not how companies choose to count contacts.

To accurately gain a truly independent view of each company's performance the quantitative measure should be removed entirely. This will result in the customers' voice alone affecting the performance league table.

Whilst the quantitative measure can be seen as a reflection of company performance, it has repeatedly been shown to be non-comparable across companies and in reality is not a true measure of customer service. Should this element of the mechanism be retained, we would strongly urge that the weighting in favour of the qualitative element be greater. Moving to option 2, with the balance of the qualitative element at

75%, would mean the customers could be assured that they heavily influence the mechanism with the company being held to account.

We maintain that the current measure is biased towards the quantitative measure (see Appendix 2). If it is retained, this bias in how scores are combined should be addressed.

Some companies have raised concerns over the potential risk of manipulation of the qualitative results. We would support a move to remove the notice period for the satisfaction survey.

We believe the current mechanism does not provide a strong enough voice for the customer. By removing the quantitative element entirely, or as a minimum reducing its weighting to 25% and removing the bias towards the quantitative score inherent in the current calculation, the customer will be confident that they are influencing the service they receive.

**Removing the quantitative measure entirely is our preferred choice, option 3, making SIM a true outcome based measure.**

- **Q6 Design issue 4 – design of the qualitative and quantitative measures**

We support the proposal to make changes in order to reflect the separation of household/non-household price controls however would suggest that for both the quantitative and qualitative measures, non-households are excluded.

Assuming the quantitative measure is retained, the proposal to count all contact routes in the unwanted contact measure is a serious concern. There is a serious risk that should this include contacts made via social media innovation would slow, resulting in reduced contact routes for customers and increased inconsistency of reporting between companies. This would also increase the complexity of data collection on a comparable basis.

We agree that phone availability is difficult to measure consistently and we support the removal of this measure. We are of the firm belief that adverse performance in this area severely affects the customer's perception and experience and so would be captured in a company's qualitative score.

Amending the quantitative measure to count only written complaints received, would simplify the data collation and make company data more comparable.

**Suggestion:** The introduction of an Alternative Dispute Resolution (ADR) scheme provides the opportunity to remove unwanted phone contacts from the measure and replace with a weighted penalty for any contacts received by the ADR provider. This

would be an independent penalty similar to the CCW investigations and would improve the comparability of reporting.

Quantitative (if retained) being:

5 points	Written complaint
100 points	Repeat Written
1000 points	CCW investigation
5000 points	ADR provider finding in favour of the customer

CCWater (and the ADR) will require this information to be provided in a consistent manner in any case. It is our view that this provides sufficient scrutiny and incentive to improve quantitative metrics, without the need to include them within SIM, but if a quantitative element is retained, this would provide a consistent comparator to replace unwanted calls.

A shortening of the survey may encourage greater acceptance to take part by the customer. And if the questions regarding key drivers on experience and satisfaction are retained we would support this proposal.

Moving to an all contact survey regardless of resolved or not would inevitably reduce the reported satisfaction levels. The result being our industry would externally appear to have worsened when in fact the measure had changed. Customers are approached from multiple sources to partake in customer satisfaction surveys and we believe it is imperative that only resolved contacts are selected for the purposes of SIM. Should all contacts be included we risk some customers receiving multiple survey calls. This would antagonise our customers and worsen the perception of our industry.

A potential to widen the scope of the survey to include customers who have had no contact with the company should perhaps be considered as an alternative to unresolved contacts.

We would however support a move to reduce the notice period for the qualitative survey.

In summary we believe:

- non-household contacts should be removed from both the quantitative and qualitative elements
- the unwanted contact measure should not be widened to all contact routes the unwanted contact should be replaced with a weighted ADR complaint penalty
- the phone availability measure should be removed
- the qualitative survey can be shortened whilst retaining the key questions
- the survey sample should include only resolved contacts but perhaps could include non-contacted customers and the notice period should be removed.

**In principle our preferred approach is for option 2 with slight modification to remove 'all contact routes' and remove non household customers from both quantitative and qualitative reporting.**

**Q7 Design issue 5 – retain the current relative incentive or use an absolute performance level to set reward and penalty**

Whilst the margins of improvement are decreasing we cannot lose sight that over recent years there has been a tremendous improvement in service levels across our entire industry. The current relative performance structure creates a huge incentive for those companies at the top of the league table. Differences in scores are at times marginal, however we strive to achieve more and to attain the number one position.

Using an absolute performance level for any reward or penalty would not in our opinion achieve the same drive for improvement. Setting an absolute performance target would be stating that a threshold level of service is acceptable. This would be firmly against the current direction of travel at Ofwat, where companies are encouraged to satisfy their customers and stakeholders, not the regulator. The purpose of both SIM and its predecessor, OPA, has been to simulate competitive service delivery. This is lost if it is no longer a relative performance measure.

**In order to maintain momentum in service improvement we recommend that the current relative incentive structure is retained, option 1**

- **Q8 Design issue 6 – non household SIM design for Wales**

We don't believe it would be appropriate to rank our household performance against non household businesses in Wales. The distinctly differing customer types have different needs and service requirements.

**In principle the incentive should operate on relative performance, but in the absence of comparable data, an absolute level might be necessary.**

**Appendix 2**  
**Quantitative bias in the current SIM**

The combination of quantitative and qualitative scores currently used in the calculation of an overall SIM score does, in theory, give equal weight to both scores. Despite this intention, the method used to normalise the quantitative and qualitative raw scores prior to combination, creates a bias towards the quantitative element, effectively making overall performance more heavily dependent on the quantitative rather than the qualitative. This appendix examines why this is the case and the implications for how recent data should be interpreted, for consideration in any change to the existing weightings.

**A simplified example**

The causes of this bias are subtle and not immediately obvious. To illustrate the problem more clearly, we present a simplified example. Suppose a school wishes to award a prize for the best overall student across two classes: Maths and English, both with equal weight in the assessment. They have narrowed the students down to a shortlist of three students whose marks shown below:

<b>Student</b>	<b>Maths (% score)</b>	<b>English (% score)</b>
Student A	<b>90%</b>	80%
Student B	80%	<b>90%</b>
Student C	80%	80%

Student A has the best score in Maths (90% compared to 80%), but Student B is equally ahead for English (again 90% compared to 80%). Intuitively, there is nothing to choose between the students, given that both classes are to be given equal weight.

The school decides to use a method similar to the SIM to make the assessment: i.e. to normalise the raw scores (0 to 1), multiply each by 50 and then add the two scores together. If both scores are normalised across the full range of points available, as expected, the scores are indeed equal for both Student A and B:

<b>Student</b>	<b>Maths (% score)</b>	<b>English (% score)</b>	<b>Overall</b>
Student A	$(90/100)*50 = \mathbf{45}$	$(80/100)*50 = \mathbf{40}$	$45 + 40 = \mathbf{85}$
Student B	$(80/100)*50 = \mathbf{40}$	$(90/100)*50 = \mathbf{45}$	$40 + 45 = \mathbf{85}$
Student C	$(80/100)*50 = \mathbf{40}$	$(80/100)*50 = \mathbf{40}$	$40 + 40 = \mathbf{80}$
<b>Average</b>	<b>41.7</b>	<b>41.7</b>	<b>83.3%</b>

However, instead of normalising across the full range of possible scores (0 to 100%), the Maths teacher chooses to normalise across the range 50 to 100%, since no student scores lower than 50% in their class. In theory, both classes are still combined with the same weighting of 50:50, but now the scores are:

<b>Student</b>	<b>Maths (% score)</b>	<b>English (% score)</b>	<b>Overall</b>
Student A	$(90-50/100-50)*50 = \mathbf{40}$	$(80/100)*50 = \mathbf{40}$	$40 + 40 = \mathbf{80}$
Student B	$(80-50/100-50)*50 = \mathbf{30}$	$(90/100)*50 = \mathbf{45}$	$30 + 45 = \mathbf{75}$
Student C	$(80-50/100-50)*50 = \mathbf{30}$	$(80/100)*50 = \mathbf{40}$	$30 + 40 = \mathbf{70}$
<b>Average</b>	<b>33.3</b>	<b>41.7</b>	<b>75%</b>

Firstly, this leads to a lower overall score for all students, as each % point lost for Maths now counts twice as much as before. But, since Student A loses fewer points for Maths than Student B, Student A's combined score now exceeds that of Student B.

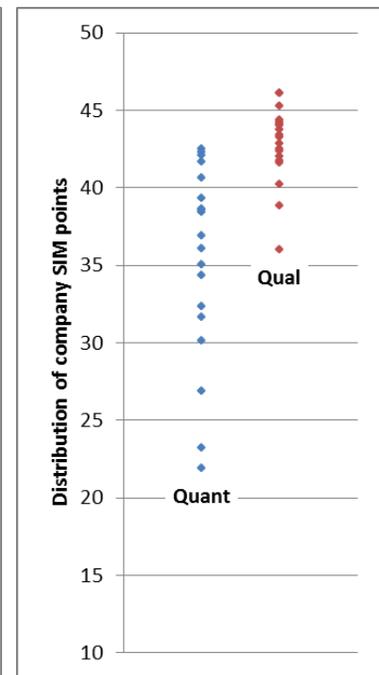
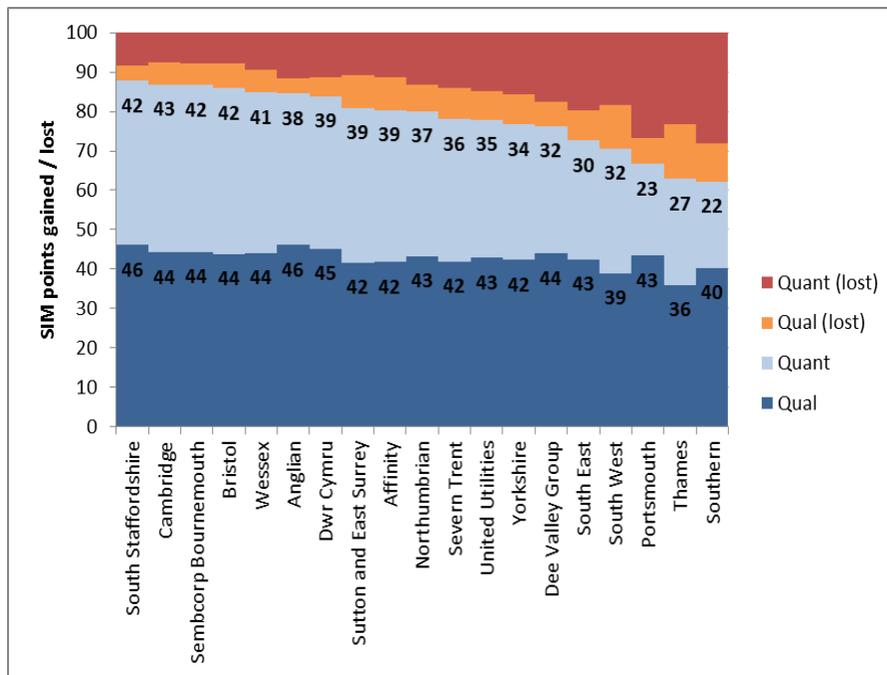
Looking at these final scores for all students it appears that overall scores for Maths are lower than for English (avg. of 33.3 vs 41.7), when in fact the distribution of raw scores for both classes is exactly equal. It also appears that Student A is clearly better than Student B (80 vs 75).

Whilst student B has some cause to feel aggrieved that they have unfairly missed out on a prize, it is of more concern if the school were to base policy decisions on this outcome: e.g. increased emphasis on Maths to make up for the apparently lower performance.

The rest of this appendix shows that a similar bias exists in the calculation of the SIM, despite the intention to give equal weight to both quantitative and qualitative elements.

## SIM: Where points are gained and lost

Figure 1 shows where companies gained or lost SIM points in 2012-13. There is comparatively little differentiation between the number of SIM points lost to the qualitative element, with all companies achieving between 36 and 46 points (a range of 10). For the quantitative score, there is much wider variation, with scores of between 22 and 42 (a range of 20). This pattern of distribution is highlighted in Figure 2, and this analysis shows that the quantitative element accounts for twice as much of the variation in company scores as the qualitative element.



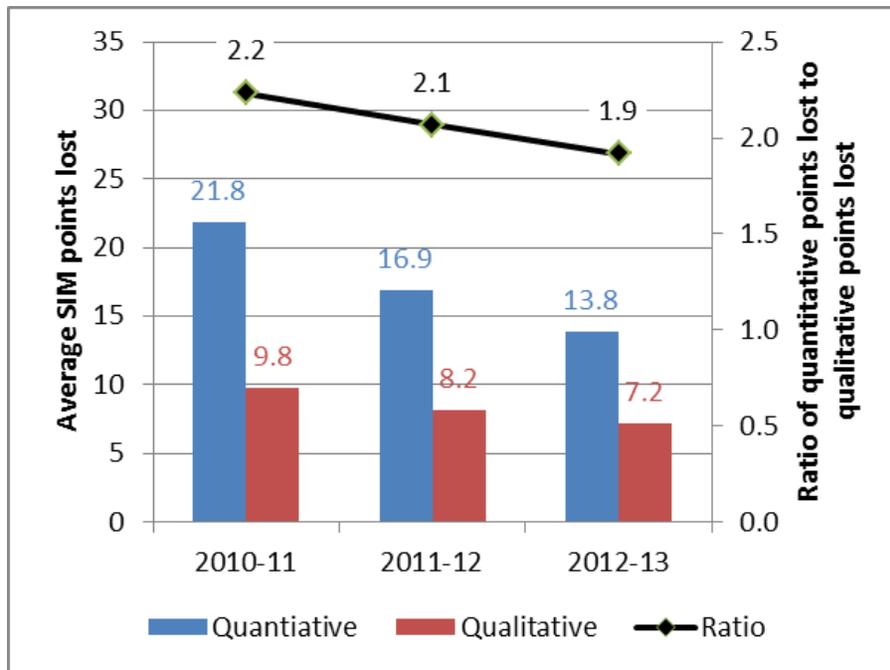
**Figure 1 SIM points gained and lost (2012-13)**

**Figure 2 Distribution of SIM scores**

On average, companies have lost 14.6% of their SIM points from the quantitative scores and only 7.2% from the qualitative element in 2012-13 – again this is around twice as much.

It might be argued that companies have improved the qualitative scores more quickly than the quantitative scores, and that this accounts for the difference in points lost.

We have previously shown that a similar pattern existed for 2011-12, and in fact, the ratio of points lost from quantitative to qualitative has actually *reduced* when assessed out of 50 (see Figure 3).



**Figure 3 Ratio of points lost from quantitative to qualitative elements**

Since its introduction, there has been significantly greater scope for improving SIM scores through the quantitative element than the qualitative.

### Reasons for this difference

The source of this bias is the ranges over which the quantitative and qualitative elements are assessed. For the qualitative measure, the full range of possible values is considered (1-5). The observed values so far do not vary greatly along this range, being concentrated at the upper end. However, for the quantitative measure, the range is set according to the observed values during testing phases (0-600). The possible range is theoretically infinite. However, whilst it is possible but unlikely that a company might have a quantitative score in excess of the range 0 to 600 (a major call centre failure for example would be one eventuality), it is *extremely* unlikely that a company would get a SIM score of 1 (every customer surveyed in a year would have to have been very dissatisfied).

The result is that much greater variability along the range of values included in SIM is seen on the quantitative measure than the qualitative. Figure 4 shows this variability for 2012-13.

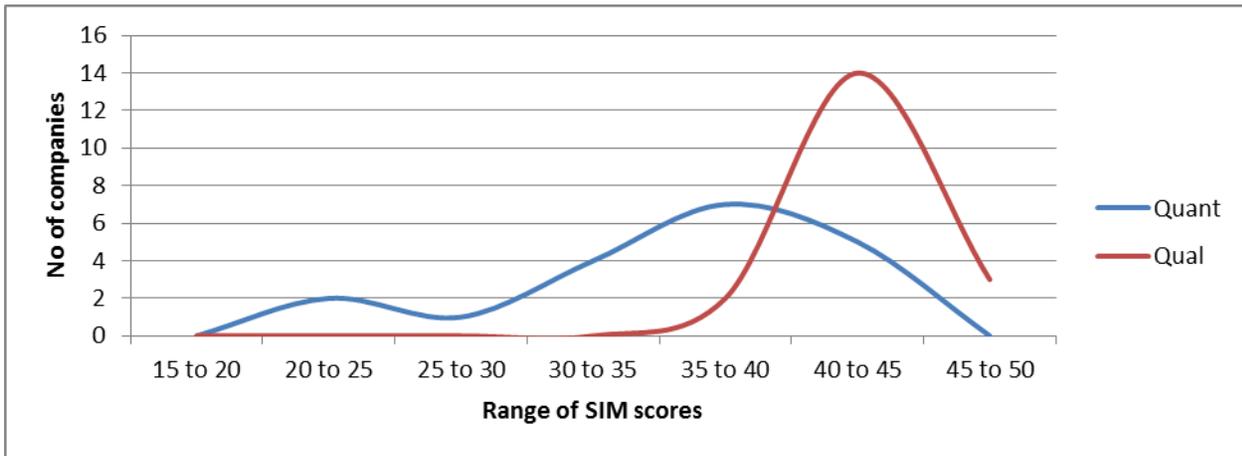
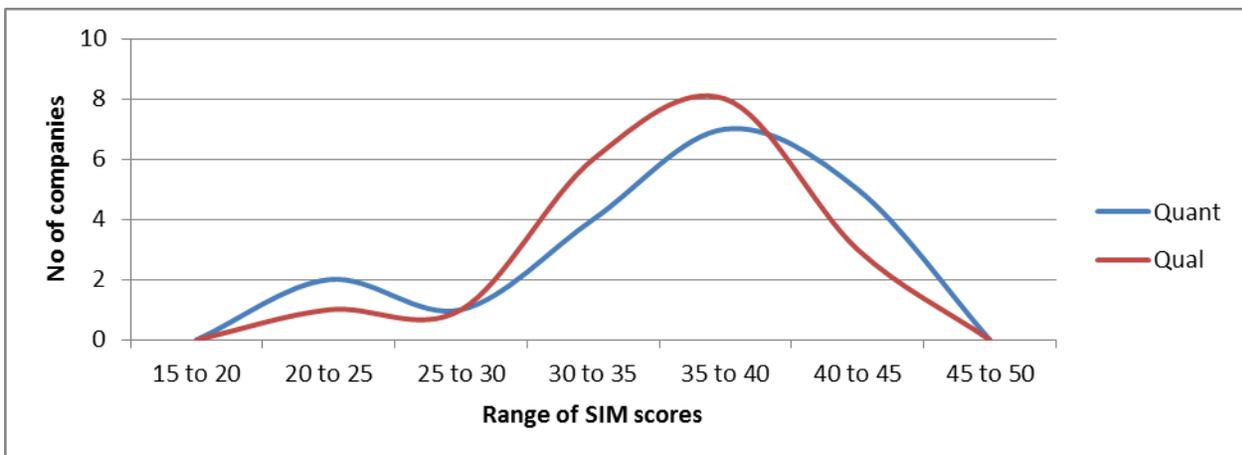


Figure 4 Range of SIM scores 2012-13

This variability is an artefact of the way the raw SIM scores are normalised to form a score out of 50.

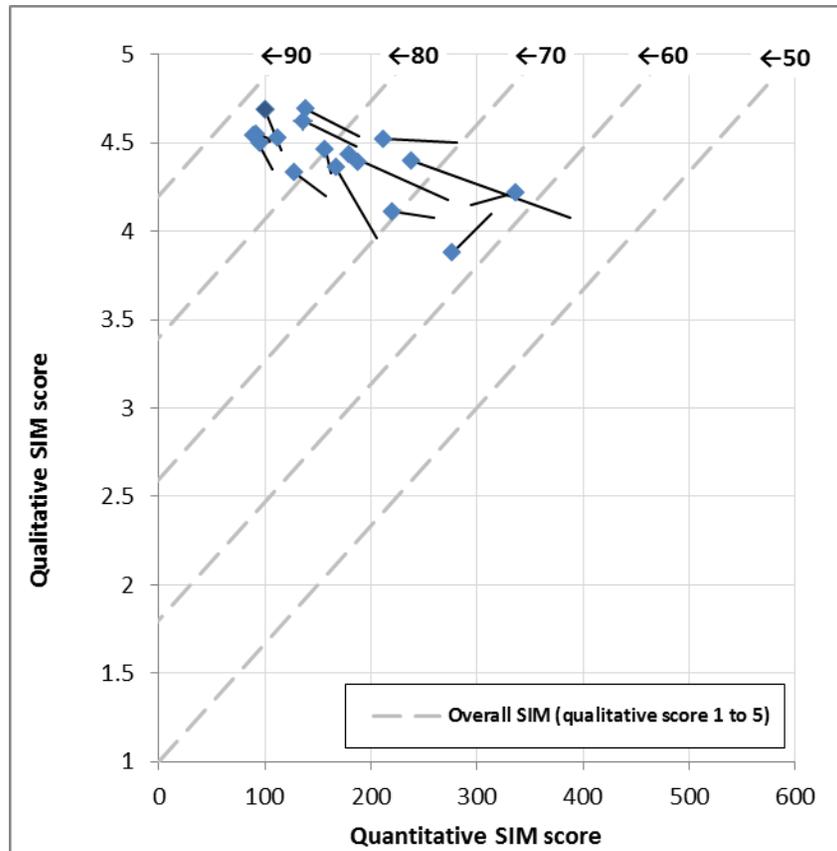
Figure 5 shows how the distribution of scores would look if the qualitative element were normalised using a range of 3 to 5 (easily covering the observed values) compared to the existing range of 1 to 5.



**Figure 5 Range of SIM scores using a qualitative range of 3 to 5**

## A biased view of industry performance

Figure 6 shows movements in quantitative and qualitative SIM scores from 2011-12 to 2012-13. The axis are set to equal the range over which each element is normalised, so that movements along the diagonal from bottom right to top left would indicate broadly equal improvements in SIM points in the final calculation. The hashed diagonal lines show overall SIM scores associated with different points on the graph.



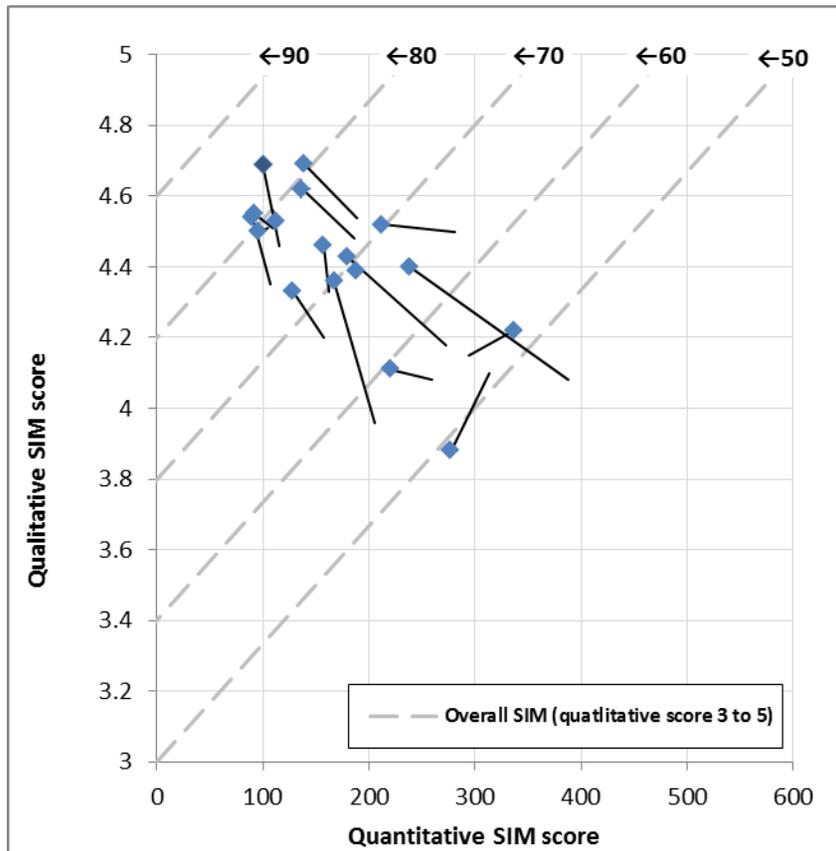
**Figure 6 Company SIM 12-13 and movement from 11-12 for quantitative and qualitative elements**

There is a tight bunching of qualitative scores in the 3.5 to 5 range (37.5% of the available range). No-one has ever scored less than 3. Quantitative scores are more widely distributed, between 0 to 400 (66% of the available range).

In general, most companies have improved both the quantitative and qualitative scores, but it appears as if improvements have been relatively larger on the quantitative scores (greater horizontal movement than vertical).

### And with the bias removed...

By changing the range over which the qualitative score is normalised to between 3 and 5 (rather than 1 and 5), the same chart shows a different picture. Using this alternative range, and again using the same axis on the chart as used in the normalisation, the bunching of qualitative scores has disappeared.



**Figure 7 Company SIM 12-13 and movement from 11-12 for quantitative and qualitative elements. Qualitative scores normalised using a range of 3 to 5.**

Now, improvements in both elements appear similar in magnitude (broadly along the bottom right to top left diagonal). This shows that apparent changes in the overall score are heavily dependent on the range used to normalise individual elements.

Substantially increasing the quantitative range would also have a similar effect.

## **Why this matters**

The effect of this bias is that the current SIM is not weighted 50:50 in practice. There are bigger gains from quantitative improvement than from qualitative improvement. Companies wishing to catch-up to the frontier may therefore be lead to heavily focus on quantitative improvement, rather than the customer experience.

This causes significant issues when interpreting the effect of the SIM:

- companies appear to be worse at quantitative than qualitative elements
- SIM appears to be driving quantitative improvement over qualitative improvement (i.e. not just the score, but the underlying factors)
- the difference in customer experience between companies is underplayed
- the difference in quantitative service measures is overplayed.

It also means that any move towards a more qualitative SIM (e.g. 75%) would actually be a much lower shift towards the qualitative element in practice

## **Recommendation**

If it is Ofwat's intention that the weighting applied to the calculation (e.g. 50:50 or 25:75) should be free from such opaque bias, we recommend that the qualitative element of SIM should be normalised over a range of 3 to 5, rather than 1 to 5. This would mean that qualitative and quantitative scores can be compared, and un-biased inferences can be made as to the effectiveness of SIM on both aspects of service.