People Driving Change

Data Accuracy and Target Setting - Stakeholder Research

Scottish Hydro Electric Transmission Limited

Created by: Matt Davis
Date Issued: 07/02/2013
Version: 1

Contents

Outline

Aims of the document

The principal aims of this document are to assess the appropriateness and accuracy of measurements in the quantitative data achieved for the stakeholder research project. The size and variability of our sample are compared with industry recognised requirements to gauge confidence in the key messages from our research.

The assessment of accuracy will be used to guide decisions on appropriate levels of performance for future research within this setting. Direction on the suitability of target levels that have already been discussed is included.

The vast majority of this document will concern the quantitative data set, since this is the sample on which our reporting is based. We may at times refer to equivalent features of the prior qualitative stage, if this provides some benefit to our evaluation. Though, this will be made clear and differences in the interviewing structure and method of capture for this data should be kept in mind.

Key considerations

The data we have collected and the population from which we have sampled are considered to be relatively small for the types of statistical analysis that have been conducted. As such, we would recommend that these findings be considered as indicative rather than fact. Statements we make within the document will relate to what is expected, based on the findings from data we have achieved. It is always possible that future trends may deviate from what is expected.

Levels of variation and accuracy have been based around the overall engagement measure Q16 (On a scale of 1-10, where 1 is low and 10 is high; please rate how satisfied you are with your overall engagement with SHETL?). This question has been selected as it is the overall measure of success in the key theme we are investigating. However, variability and the associated accuracy are specific to each question and as such; other measures may have considerably differing levels of accuracy.

Distribution of responses

Overall engagement responses

The response distribution for the overall engagement measure is typical when looking at survey distribution. The vast majority of responses were positive and therefore a somewhat positively skewed response distribution is formed.

The minimum option that a respondent could select was a score of 1 out of 10 and the maximum a 10, however, the lowest score received was 2 out of 10 (from 1 respondent).

By far the vast majority of responses we received considered engagement to be rated either 7 or 8 out of 10. 63% of respondents gave a 7 or 8 rating to the engagement measure meaning this is where the majority consider current performance to be.

All measures of central tendency were between values of 7 and 8. We computed a mean score of 7.2 for our sample and both the median and mode were a rating of exactly 8 out of 10.

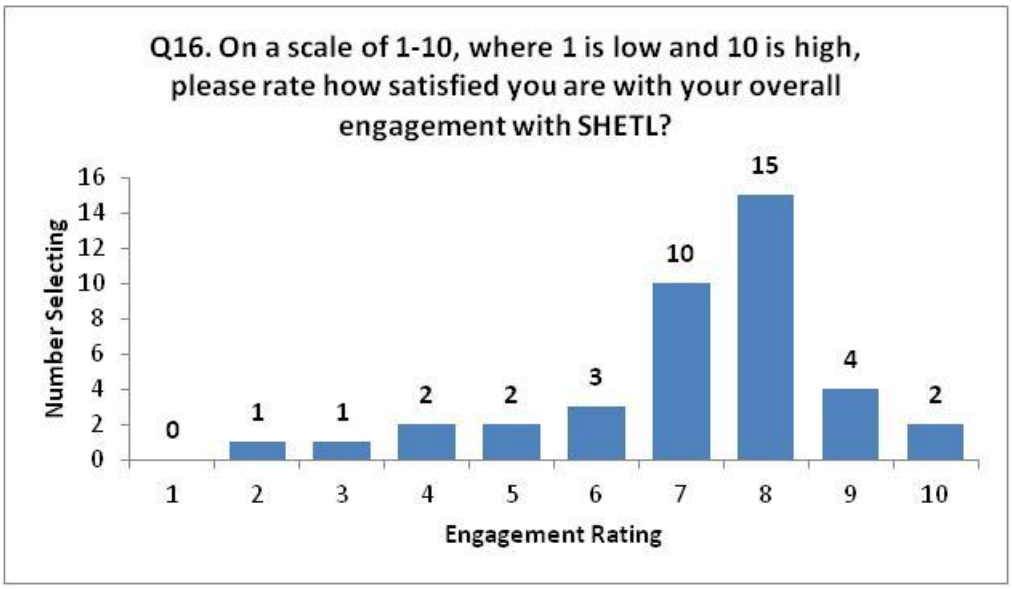

Fig 1. Overall engagement distribution plot

The qualitative equivalent

The rating of overall engagement was also asked to respondents during the qualitative stage of research. 15 responses were received that gave a rating between 1 and 10. The equivalent mean score for this sample was 6.7, with median and mode values of 7 and 8 respectively.

Interestingly, there was a respondent in the 15 qualitative interviews that gave an extremely low score of 1 out of 10 (with the next lowest being a rating of 6). If we remove this respondent as an outlier, results in the qualitative data become very similar to the quantitative stage. The mean score for this sample would be 7.2 and although the median and mode are still 7 and 8 respectively, the percentage of 7 and 8 responses becomes 64%. It may be a coincidence but considering the size of the samples involved, it appears that the distribution and location of the quantitative and qualitative data we have collected are analogous.

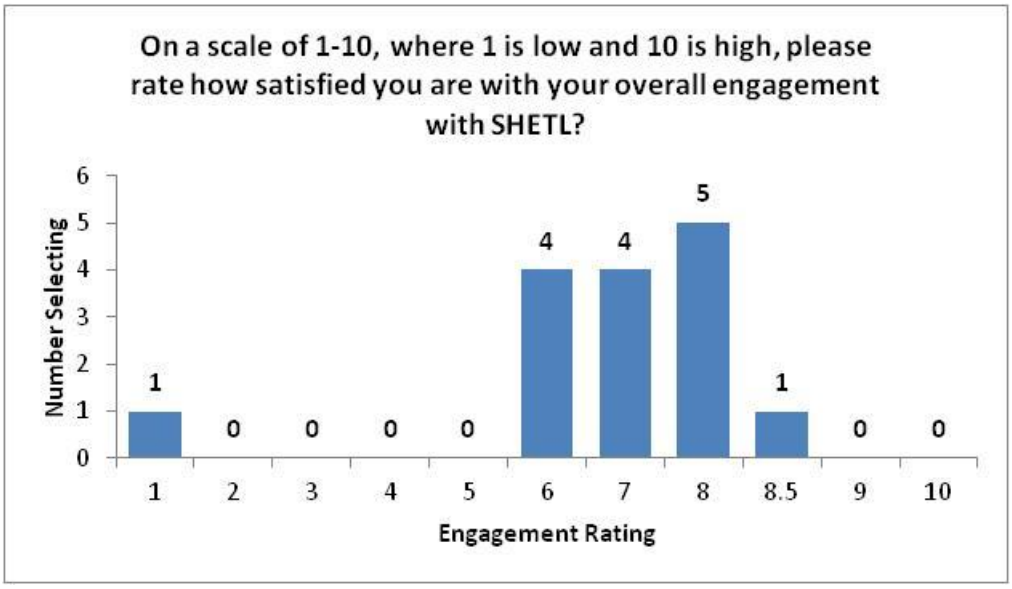

Fig 2. Overall engagement distribution from the QUALITATIVE stage

A representative sample

A key area for concern when assessing the quality of the sample we have achieved is whether it is representative of the population we are describing. As we have a total sample base of 40 respondents, the sub-groups in our data are often extremely small. However, one possible concern is that the lowest ratings for overall engagement have come from a particular sub-group in Connections/Construction. Only 4 interviews from this group were achieved, but their ratings were very poor. If the group were under-represented in our sample then future research, where the sample included higher numbers from this group, could potentially have much lower scores.

As the overall sample we collected contained 40 respondents, 4 of which were Connections/Construction, they represent 10% of the sample. For the entire database the equivalent figure for the proportion of Connections/Construction was found to be 9%. Which suggests our sample is representative for this group and we would not expect this to impact on future scoring.

Initially suggested target

A rating of 5 out of 10

The very initial discussions surrounding the setting of an appropriate target have mentioned a rating of 5 out of 10 as conceivably being the accepted level of performance to achieve. Even with the small amount of evidence we have seen, it is clear that this rating is low when compared with the vast majority of responses we have received. In the illustration below, for a given target level, the percentage of respondents who would score above the target, at the target and below the target is shown. The percentages are based on the quantitative sample we have received.

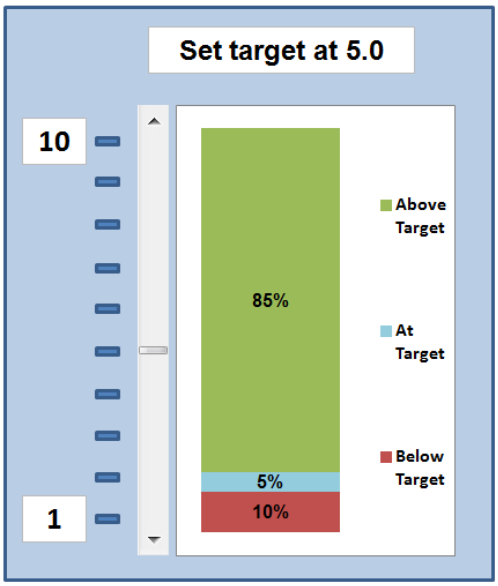

Fig 3. Likely performance against a target rating of 5

In the data we have received, 85% of responses scored above a rating of 5 out of 10. Also, 5% of responses were a rating of exactly 5 out of 10. So if the target performance were set at this level, 90% of the responses we have received are at or above this base level.

We can be fairly certain that the mean score we compute from these responses will be above this level, since so few ratings fall below the target. For the next wave of research to score anywhere near this level, we would need to see a very extreme change from the current level of performance.

Future waves of research

We have mentioned that to be scoring at a level of around 5 out of 10 we would need to see an extreme change from the current rating. It is of course possible that this will be the case and significant movements from one wave of reporting to the next are common, however, movements as extreme as those required here are more exceptional.

It has also been suggested that the scoring in subsequent waves of research often decline when compared with the first wave of interviewing. In our experience the opposite is more commonplace, as the level of engagement most often shows an incremental increase wave on wave as service levels are monitored and improved with the assistance of the research programme. A reduction in the magnitude of such improvements wave on wave is familiar as it becomes harder to improve from a higher starting point.

Statistical confidence in our sample

Variability in our sample

The quantitative sample achieved a mean score of 7.2 from 40 interviews. The standard deviation, or average distance, from this mean was 1.7. Another way of thinking about this is standard deviation is that "on average" our scores are between 5.4 and 8.9.

Sampling distribution

This is however, simply a measure of the location and variability within the sample we have collected. It is perhaps more useful to know how successful this sample is at describing the overall population of stakeholders.

We can use the size and variability of our sample to make inferences on estimate reliability for the population. We can also incorporate the mean of our sample and the standard error of the mean, to give an estimate of the true population mean.

With an overall population of 225 to describe, and given that our sample suggests the standard deviation of responses is 1.7, the sample size requirement for industry recognised levels of confidence (95%) and accuracy (5%) is 46 respondents. This requirement is based on reporting mean score figures, such as our sample mean of 7.2, and it includes a correction factor applied due to the low population figure.

What does this mean?

We have stated that for a population of 225 we require 46 respondents to report mean scores with 95% confidence and 5% accuracy. In terms of our sample, this does not mean a lot, other than we have only 40 interviews so we have fallen short of the requirement. In reality it means a slightly 'wider' estimate for the true mean of the stakeholder population as we make statements like those below:

- *We can be 80% confident that the actual engagement rating for the whole population of stakeholders lies in the interval 6.9 to 7.5*
- *We can be 90% confident that the actual engagement rating for the whole population of stakeholders lies in the interval 6.8 to 7.6*
- *We can be 95% confident that the actual engagement rating for the whole population of stakeholders lies in the interval 6.7 to 7.7*
- *We can be 99% confident that the actual engagement rating for the whole population of stakeholders lies in the interval 6.5 to 7.8*

So even though we have spoken to only 40 of 225 stakeholders, we feel almost certain that the current level of engagement for all stakeholders is between 6.5 and 7.8 out of 10. These estimates, however, rely on the variability in our sample of 40. It is important to remember this, as we may find that a larger sample is in fact more variable. In essence, we can trick ourselves into thinking our data is more robust than it is as the small number we spoke to, answer in a similar way. We would account for this by perhaps allowing more leeway in our estimation of limits, for example using the 99% confidence limits above in preference to the 95% equivalent.

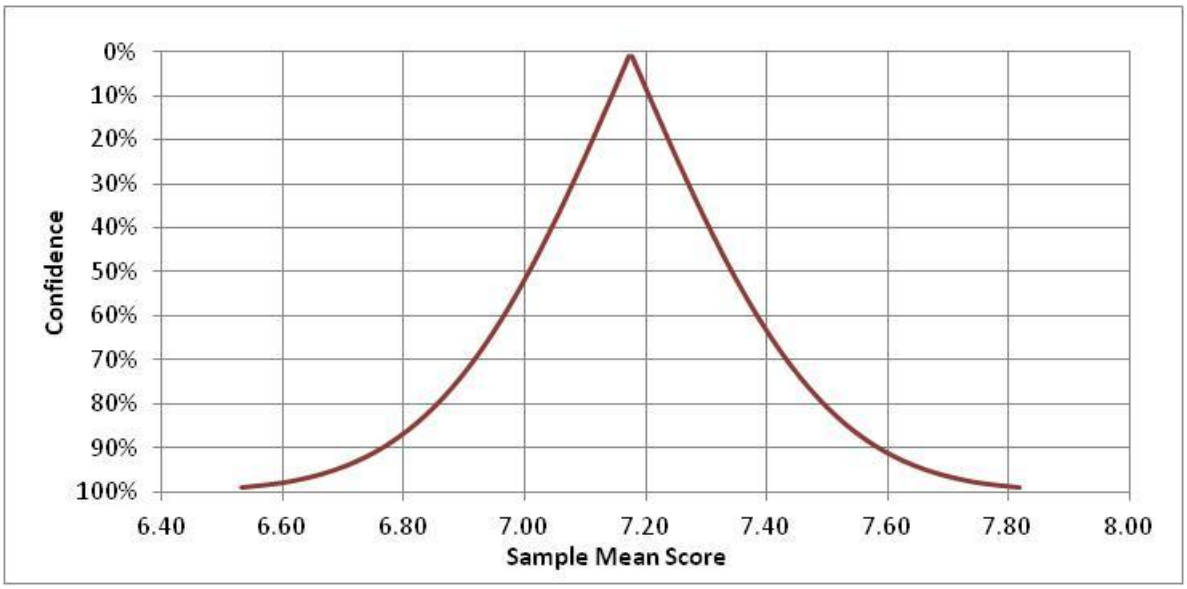

Fig 4. The approximate sampling distribution of means

It is clear how far the target of 5 out of 10 is from this level and therefore, given the current level of service, how unlikely we are to select a sample that has a mean score of 5.

A change in performance

It is of course true that performance can change rapidly and the engagement score would change with it. We know that setting a target without the knowledge of current business plans and several waves of data to trend may be short-sighted. Particularly if the target we set is long term and cannot be revisited. However, it is our opinion, that a mean score as low as 5 in subsequent waves of research is extremely unlikely. It represents a significant shift much more extreme than we are used to seeing.

To illustrate this, the change in engagement mean score required to constitute a statistically significant movement up or has been calculated. This is based on the current sample size and variability and assumes that these parameters will remain consistent in subsequent waves of reporting.

- *For a significant rise in engagement score from our current rating of 7.2, the score would need to move to 7.9.*
- *For a significant fall in engagement score from our current rating of 7.2, the score would need to move to 6.4.*

A target level of 5 is well outside of this range and applying the same assumptions it represents what would be several statistically significant shifts down in the current high levels of engagement (i.e. it is just under 3 times the movement required to see a statistically significant score change).

Suggested Target

Summary

When suggesting a target we would normally conduct analysis to determine current trends, performance boundaries and the suitability of the target performance level for all sub-groups of interest. This would be combined with knowledge of business objectives and industry concerns to set the most appropriate target level. As we have conducted only one wave of research here, trending and forecasting is not possible and cannot assist our analysis.

Any recommendation for the setting of a performance target is therefore limited to the data we have collected and information relating to industry movements and the processes occurring within SHETL itself.

The data suggests that the current level of engagement is between 6.7 and 7.7 (this is based on a 95% confidence interval on our mean score of 7.2). This represents very positive performance on a scale of 1 to 10. If industry movements are expected to be small and the processes occurring within SHETL are not set to change dramatically, then we would recommend an internal engagement performance target within these limits.

Moreover, for the performance of another wave of interviewing (with similar variability and sample size) to be a significant movement, the engagement score would likely need to be outside of the range 6.4 to 7.9. We would therefore suggest that unless significant changes are happening, an engagement target outside of this range would result in performance a long way above or below the target level. We would consider these bounds to be the lower and upper limit of expected performance and that 6.4 perhaps best represents the base performance level we would aim for internally (i.e. this target represents avoiding a statistically significant fall in engagement score). It is worth remembering that the current performance and the vast majority of responses already received are above this level.

Benchmark for Stakeholder Satisfaction Incentive

This information relates to an internal engagement target, based on continuing the current high levels of engagement seen in our survey at Q16. When extending this to think about an Ofgem benchmark performance level of 5 out of 10, we must remember that the current performance perhaps should not represent the base or mid-point on the performance scale.

We have only contacted a relatively small sample of 40 respondents which, as mentioned previously, therefore we should conceivably be prepared for scores to move further than if the sample were larger.

We quoted a range relating to current performance, where we are 99% confident our score is above 6.5 and that a similar score of 6.4 is the point at which performance would be on the border of a significant fall. If we trust that the sample we received is completely representative and that fluctuations would be this small, we would recommend that this is close to the approximate mid-point level and that a benchmark of 6, subtracted from the current stakeholder satisfaction score (overall satisfaction with engagement measure), may well result in the most appropriate adjustment factor.

A benchmark of 5 represents a more conservative level for the data we have seen. Although we are likely to achieve a sample of similar size in subsequent years, it is our experience that the variability estimates in small samples can be misleading and the performance level may change more dramatically than conventional statistics suggest. It is certainly true that this variation works both ways but we would emphasise that the data supporting an increase to the benchmark, potentially suffer from this concern. In essence, the evidence to suggest we should increase the benchmark from a score of 5 to 6 is less dependable than we would ideally like, when making such recommendations.