~~Chapter 3:~~Article 3 of 6

~~q~~Questionnaire D~~d~~esign

The customer satisfaction measurement requirements of ISO 9001:2000 will lead to an increase in the use of surveys. The best response rates will be achieved by the most professional questionnaires. **Nigel Hill**, of the Leadership Factor, considers questionnaire design, wording and rating scales, and highlights some serious mistakes often seen on questionnaires

The standard says nothing specific on the technicalities of questionnaire design. The only clues to designing a questionnaire are to be found in the 'monitoring and measurement' section, where clause 8.2.1, 'customer satisfaction', states: 'As one of the measurements of QMS performance, the organisation shall monitor information relating to customer perception as to whether the organisation has met customer requirements. The methods for obtaining and using this information shall be determined'.

The organisation must measure whether it 'has met customers' requirements': this has two implications for questionnaire design. First, customer requirements must be established through exploratory research using the 'lens of the customer' (as explained in chapter two of this series of articles). Second, the questionnaire must cover importance and satisfaction - the former for the relative importance of customers' requirements and the latter for the extent to which requirements have been met.

Questionnaire design

Questionnaire length
Questionnaire design begins with customer requirements identified by exploratory research. Fifty questions are the maximum number recommended; the number of requirements included therefore depends on the number of additional questions. Most additional questions are used to classify respondents and some questions may have nothing to do with customer satisfaction. It will take 40 questions to cover the top 20 customer requirements (identified from exploratory research), since each must be rated for importance and satisfaction. Probing low satisfaction scores generates an average of three additional questions per respondent. Four classification questions would leave three additional questions for other topics. Figure 1 shows questionnaire sections.

Fifty questions will take up to fifteen minutes to answer, whether the method is interview or self-completion. With paper questionnaires, fifty questions can be squeezed onto a double-sided questionnaire or can be spread over four sides. Although shorter questionnaires are desirable per se, the four-sided questionnaire will achieve a higher response rate and a better quality of response because it looks more attractive and is easier to complete.

Instructions

Instructions will be seen first on self-completion questionnaires, so must be crystal clear. If a questionnaire can be filled in incorrectly, it will be, so thorough instructions are essential, even if they are lengthy.

Satisfaction and importance questions

There must be separate sections for satisfaction and importance. Covering importance and satisfaction for each requirement at the same time produces an artificial correlation between the two. Although it is conventional to place importance before satisfaction, tests by The Leadership Factor show that it is better to start with satisfaction. This way respondents are familiarised with the issues before they are asked to score requirements for importance. When importance follows the satisfaction section, scores for importance vary much more, providing greater discriminatory power for analysis. Satisfaction scores remain similar in spite of the section's placement.

Classification questions

Classification questions appear last. Some people see questions about age, occupation or income as impertinent, and may abort the interview or questionnaire if they appear at the beginning.

Questionnaire wording

The two main wording problems identified on satisfaction questionnaires are due to ambiguity and bias. Both lead to flawed results.

Ambiguous questions

Each question and its corresponding answer must have one clear meaning. Pitfalls such as jargon and acronyms should be obvious, but there are more serious problems such as double questions, which lead to misunderstanding and non-actionable results. One example is: 'Were the staff friendly and helpful?' Which characteristic do you want to measure? Friendliness and helpfulness are not the same. If that question scores poorly, which area would you target for improvement? To measure both of these aspects of staff behaviour, you must ask two questions.

Biased questions

Your questionnaire can distort results through biased questions or a biased rating scale, for example:

- how satisfied are you with the variety of food on the menu?
- how satisfied are you with the speed of response for on-site technical support?

The questions suggest that customers are already satisfied and asks them to rate how satisfied they are. An unbiased version of these questions would be:

- how satisfied or dissatisfied are you with the variety of food on the menu?
- how satisfied or dissatisfied are you with the speed of response for on-site technical support?

Rating scales

Biased rating scales

The rating scale can also bias the response. Biased rating scales are common (see figure 2). The scale shown in figure 2 is not balanced and will bias results towards satisfaction. Most positively biased rating scales on customer satisfaction questionnaires are due to poor questionnaire design. However, some organisations that are very experienced in CSM deliberately use positively-biased questionnaires on the grounds that only 'top box' satisfaction matters, so only degrees of satisfaction are worth measuring. There are two problems with this philosophy. Even if most customers are satisfied, it is still essential to understand customer dissatisfaction. It is more valuable to the organisation to identify problem areas than to have detailed information on the satisfaction levels of its most satisfied customers. Second, positively biased rating scales are not necessary. Wider scales can adequately accommodate satisfaction and dissatisfaction.

Figure 3 illustrates a balanced rating scale because it has an equal number of points above and below the mid-point and the words at opposing points on the scale are exact opposites. It would be balanced with or without a mid-point. It is a myth that many respondents choose the middle option. They say what they think and produce a range of answers. Some respondents avoid the extremes of the scale, which explains clusters around the mid-point. This creates a problem with a five-point scale because respondents avoiding extremes have only three points left.

Since the rating scale is the satisfaction measurement tool, its choice is critical. The three most commonly used scales in satisfaction research are Likert, verbal and numerical scales, shown in figures 4 to 6.

Likert scale

Very common in attitude research, the Likert scale is easy to fill in but the bold statement introduces bias. Likert scales on satisfaction questionnaires are always positively biased. You never see questions like: 'The restaurant was filthy - agree/ disagree.'

Verbal scale

Verbal scales are easy and have the advantage of incorporating the concepts of importance and satisfaction into the scale, thus reducing respondent confusion. Reporting verbal scales on the basis of 'percentage satisfied' (ie those ticking the top two boxes) masks changes in customer satisfaction caused by the mix of scores in the satisfied or dissatisfied categories.

It is not statistically acceptable to convert the points on a verbal scale into numbers to generate an average satisfaction score because verbal scales are ordinal. In other words, we know that 'very satisfied' is better than 'satisfied' but not by how much. Nor do we know if the distance between 'very satisfied' and 'satisfied' is the same as the distance between 'satisfied' and 'neither satisfied nor dissatisfied'. Therefore, verbal scales have to be analysed using frequency distributions, which show how many respondents ticked each box. It is not statistically acceptable to use means and standard deviations, which makes it impossible to compare importance and satisfaction directly, unless points on the scale are grouped, eg the percentage of respondents ticking 'very satisfied' and 'satisfied' versus the percentage ticking 'very important' and 'important'. Due to the distribution of scores within these broad categories, such results may not paint a reliable picture.

Numerical scale

Numerical scales are easy to fill in and can be used whatever the method of data collection. Verbal scales are very clumsy in telephone interviews. Numerical scales are ideal for analysing data and communicating the results to colleagues. Average scores are easy for everyone to understand and paint a very clear picture of results and their implications. They illustrate clearly the areas to address, compared with the far less graphic frequency distributions that must be used to analyse verbal scales. This is extremely important because internal feedback is where the CSM process fails in many organisations.

Numerical scales can have more points than verbal scales. The differences between satisfaction survey results from one period to the next will often be very small and a wider scale enables the respondent to be more discriminating. Scales with more points discriminate better between top and poor performers, so tend to have greater utility for management decision making and tracking. Further, it is easier to establish 'covariance' between two variables with greater dispersion (ie variance around their means). Covariance is critical to the development of robust models, such as those that identify drivers of satisfaction. These analytical arguments will be developed in a future article.

Take-home points

- ISO 9001:2000 specifies that the organisation must measure 'customer perception as to whether the organisation has met customer requirements.' This means that the questions must be based on customers' main requirements, identified by exploratory research
- don't include more than 50 questions and, if the test is to be self-completed, spread them over four sides of paper, rather than compressing them
- score satisfaction and importance in separate sections with satisfaction first
- low satisfaction scores should be probed to understand the reasons behind them
- classification questions should appear at the end of the questionnaire
- to avoid ambiguity and misinterpretation, the questionnaire wording should be precise, with no double questions
- unbalanced questions and unbalanced rating scales introduce bias
- numerical rating scales are the most useful type as they are better for analysis and feedback
- scales with more points are better than those with fewer because they create greater discriminatory power - a ten point numerical scale is recommended