Rating Scales in Customer Satisfaction Research

1. Interval versus ordinal scales

It is not unusual in satisfaction research to see simple verbal scales, where each point on the scale is given a verbal description (e.g. 'strongly agree', 'agree' or 'very satisfied', 'satisfied' etc). The problem is that such scales have only ordinal properties. They give an order from good to bad or satisfied to dissatisfied without quantifying it. In other words, we know that 'strongly agree' is better than 'agree' but we don't know by how much. Nor do we know if the distance between 'strongly agree' and 'agree' is the same as the distance between 'agree' and 'neither agree nor disagree'. Therefore, verbal scales have to be analysed using a frequency distribution, which simply involves counting how many respondents ticked each box. It is not statistically acceptable to use means and standard deviations, to develop indices or to apply most multivariate statistical techniques to establish the relationships between variables in the data set - a major weakness for understanding things like the drivers of satisfaction and loyalty.

Interval scales use numbers to distinguish the points on the scale. They are suitable for most statistical techniques such as means, standard deviations, indices and correlations because they do permit valid inferences concerning the distance between the scale points. For example, we know that the distance between points 1 and 2 is the same as that between points 3 and 4, 4 and 5 etc, so it is reasonable to conclude that a score of 4 represents twice the magnitude of a score of 2.

For a scale to have interval properties it is important that only the end points are labelled; the labels (e.g. Very satisfied.......Very dissatisfied) simply serving as anchors to denote which end of the scale is good / bad, agree / disagree etc.

2. The meaning of words

People often subjectively prefer verbal scales because they feel that they understand the meaning of each itemised point on the scale, and at the individual

level it is true that each person will assign a meaning that they understand to each point on the scale. By contrast, some people say that a numerical score doesn't appear to have a specific meaning - does one person's score of 7/10 refer to the same level of performance as another person's score of 7. Interestingly, tests have shown that collectively, people interpret words in a wider variety of ways than numbers. One person's 'satisfied' is not necessarily the same as another's.

For international businesses the problem of consistent interpretation of verbal scales is hugely exacerbated by language and cultural differences. Any company conducting international research would be extremely unwise to consider anything other than a numerical rating scale.

3. Aggregating data from verbal scales

Since it is not statistically acceptable to convert the points on a verbal scale into numbers and generate a mean score from those numbers, the only statistically valid method of analysing verbal scales is a frequency distribution. This leads organisations to report verbal scales on the basis of "percentage satisfied" (i.e. those ticking the boxes above the mid point). This often masks changes in customer satisfaction caused by the mix of scores within the 'satisfied' and 'dissatisfied' categories. In fact, if results are reported in this way there is little point having more than 2 points on the scale - satisfied and dissatisfied.

4. Number of points

It is not practical to have many points on a verbal scale. Realistically, 7 points is the maximum. This is a considerable disadvantage since the differences between satisfaction survey results from one period to the next will often be very small. A wider scale enables the respondent to be more discriminating especially at the satisfied end of the scale, which is important since it is only the very satisfied who are likely to recommend and to remain loyal.

Interval scales most commonly have 5, 7 or 10 points. More points yield greater variability, which is better for analytical purposes for two main reasons. First, scales with more points discriminate better between top and poor performers so tend to have greater utility for management decision making and tracking. Second, it is easier to establish 'covariance' between two variables with greater dispersion (i.e. variance around their means). Covariance is critical to the development of robust multivariate dependence models such as identifying the drivers of employee commitment or customer loyalty, or establishing the relationship between employee satisfaction and customer satisfaction.

5. The danger of over-stating customer satisfaction

Due to the 2 problems outlined above (narrower distribution of scores and aggregation of data), verbal scales almost invariably generate higher customer satisfaction scores than numerical scales, tempting organisations to adopt a dangerous level of complacency regarding their success in satisfying customers.

From time to time, The Leadership Factor has to convert an aggregated customer satisfaction index from a verbal scale into the weighted *Satisfaction Index*™ used by The Leadership Factor and generated by

a numerical scale. This is done by duplicating some questions using both scales in the same questionnaire or in separate surveys conducted simultaneously. Sample data below from 5 example tests shows how results from the two different scales compared:

|	% "satisfied" 5 point verbal scale	Satisfaction Index™ 10 point numerical scale
Example 1	84.5%	67.4%
Example 2	78.6%	65.3%
Example 3	87.6%	70.9%
Example 4	90.3%	74.4%
Example 5	92.3%	75.8%

This can lead to a dangerous level of complacency. In Example 5, the 92.3% produced by the 5 point verbal scale suggests that the company is doing well at satisfying its customers. In fact, plugging its Satisfaction Index™ of 75.8% into The Leadership Factor's benchmarking database demonstrated that it was in the bottom half of the league table in its ability to satisfy customers! It is therefore hardly surprising that companies misleading themselves with unrealistically high levels of customer satisfaction from verbal scales complain that their 'satisfied' customers are often defecting. They then begin to question the point of customer satisfaction. What they should be questioning is their customer satisfaction measurement process. It's giving them a bum steer! Their customers are actually well below the levels of satisfaction that would generate loyalty.

This was demonstrated by AT&T who, using verbal scales were regularly getting CSIs of 90%+ and bonusing staff on it, but in 1997 had doubts when some businesses began making major losses despite these apparently high levels of customer satisfaction. On investigation, they discovered that repeat purchase rates were substantially different for customers rating "excellent" compared with those rating "good". They also found that CSI scores of 95%+ were correlated with scores in the low 80s on their measure of "worth what paid for". (Source, Customer Service Management magazine case study).

This is supported from a huge amount of

evidence collected over a 30 year period by Harvard Business School. They have found very strong correlations between customer satisfaction and loyalty, but only at high levels of satisfaction. Merely being satisfied isn't enough in today's competitive markets and a tougher measure based on a 10 point numerical scale is necessary to highlight this.
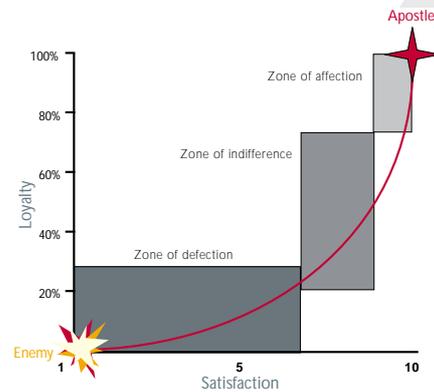
6. Academic research

For background reading, a widely respected general text that covers the subject of scales is:
J.C. Nunally, "Psychometric Theory" (1978, pages 12-20), McGraw Hill.

A more recent and more relevant treatment of the subject is found in:
D.Wittink and L.Bayer, "The Measurement Imperative". Marketing Research, Volume 6, Part 4, pages 14-22, 1994.

Wittink and Bayer took into account respondent issues such as simplicity and understandability as well as statistical considerations such as reliability (i.e. repeatability), power (detection of differences between groups or over time) and sensitivity (utility for improving customer satisfaction). Considering all factors they concluded that 10 point numerical rating scales are best for

measuring customer satisfaction.

At the present time the leading academic centre of excellence for satisfaction measurement is the University of Michigan Business School, which, amongst other things, is responsible for the American Customer Satisfaction Index. The ACSI uses a 10 point numerical scale. We asked Michigan why they use this scale and received the following reply:

"Our recommendations are based more on our experiences than any documented studies. But there are two articles that help directly or indirectly support the 10-point scales (see below). The first, by Ryan et al., compared 5 and 10 point scales to multi-item indices of 10-point scale. (Don't be deceived by the title.) The second shows how satisfaction and quality data distributions have a generalisable negative skewness (i.e., most of the observations are toward the high end of the scale). This is why, to me, it makes sense to have more than just 7 scale points.
I hope this helps!
Yours,
Michael"

References:
Ryan, Michael J., Thomas Buzas, and Venkatram Ramaswamy (1995), "Making Customer Satisfaction Measurement a Power Tool," Marketing Research, 7 (Summer), 11-16.
Fornell, Claes (1995), "The Quality of Economic Output: Empirical Generalizations About Its Distribution and Association to Market Share," Marketing Science, 14 (Summer), G203-G211."
[Personal communication from Professor Michael Johnson of University of Michigan Business School, 12 September 2001]
.

7. Distribution of data

Picking up on Professor Johnson's point, it is true that satisfaction data tends to be skewed towards the high end of the scale, although skewness is exacerbated on 5 point scales. This merely reflects the fact that most customers and employees are
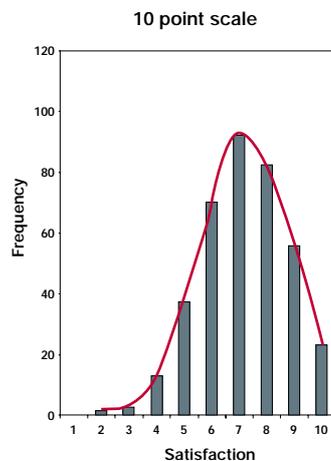
broadly satisfied rather than dissatisfied - they wouldn't still be there if they weren't. What most companies are really measuring therefore is degrees of satisfaction. (It is interesting to note that situations where high levels of dissatisfaction do exist, typically when customers have no choice, do exhibit a much more normal distribution). Given that we are mainly measuring degrees of satisfaction and tracking small changes in that zone, it becomes very important to have sufficient discrimination at the satisfied end of the scale, and, for analytical purposes, a good distribution of scores.

As illustrated by the charts below, whilst both 5 and 10 point scales exhibit a skewed distribution, data from the 10 point scale are more normally distributed and show more variance. It should also be noted that variance can
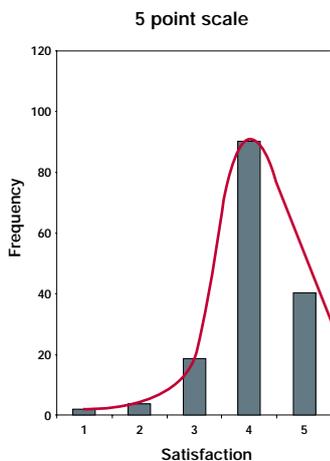
distribution with 5, 7 or 10 point scales and have little need for advanced analysis of the data since the problem areas that need addressing will be obvious. By contrast, choice of scale becomes much more critical for top performing companies for several reasons:

1. Companies with high levels of satisfaction need a very tough measure if they are to identify further opportunity for improvement.

2. Companies in this situation need to employ much more sophisticated statistical techniques that drill down into the data to uncover drivers of satisfaction or differences in satisfaction between groups of customers that may not previously have been considered.

3. In situations where there are multiple business units (e.g. branches, regions, stores, sites etc) it is very important to be able to discriminate between the better and poorer performing units.

> ANY COMPANY CONDUCTING INTERNATIONAL RESEARCH WOULD BE EXTREMELY UNWISE TO CONSIDER ANYTHING OTHER THAN A NUMERICAL SCALE.
>
> Nigel Hill,
> The Leadership Factor

measures. 20 point scales have also been used. However, questions must be easy for respondents to understand in order to have a high level of confidence in the validity of the answers. People find it most easy to respond to 5 point verbal scales and 10 point numerical scales. This may be because giving (or receiving) a score out of 10 tends to be familiar to most people - whether it be from tests at school, from the reviews of footballers in newspapers and other real life experiences. Numerical scales with fewer or more than 10 are more difficult for people as are verbal scales with more than 5 points. **S**

5 point scale

10 point scale

be affected on numerical scales by the labelling of the anchored end points. To illustrate, a 10 point scale with end points labelled 'dissatisfied' and 'satisfied' would represent a narrower range of views than end points labelled 'very dissatisfied' and 'very satisfied'. Even better would be end points labelled 'totally dissatisfied' and 'totally satisfied'.

8. Top performers and poor performers

For poor performers with low levels of satisfaction, the choice of rating scale matters little. They will get a fairly normal

For all the above reasons the greater variance yielded by 10 point numerical scales and the ability to employ advanced multivariate statistical techniques with good levels of predictive and explanatory power are extremely beneficial to high performing companies.

9. Ease of completion

In the light of all the above arguments, it would be valid to ask the question, 'Why stop at 10 points?' From a data point of view it would be better to have even more points. Federal Express uses a 100 point scale to track 'micro-movements' in customer satisfaction in its frequent

Nigel Hill
Managing Director
The Leadership Factor

www.leadershipfactor.com
email: info@leadershipfactor.com