Rating Scales for Customer Satisfaction Research

If you want to measure anything you have to have a measuring device. Unfortunately, no measuring device for customer satisfaction has ever achieved the universal adoption of the Celsius scale, the speedometer or the 12” ruler. That’s partly because there isn’t a definitive answer to the question “which is the most suitable rating scale for market research?” However, much of the reason for the proliferation of rating scales used for CSM is due to practitioners’ ignorance about the characteristics of different scales and their advantages and disadvantages for customer satisfaction measurement.

There are two key reasons why organisations conduct customer satisfaction surveys:

1) Taking action

Since organisations can’t address everything simultaneously they need to prioritise the allocation of resources. Without measures it would be impossible to make reliable decisions about the best areas to focus resources to improve customer satisfaction.

2) Judging success

If customer satisfaction is a key indicator of business performance, trying to improve it without a yardstick for judging success would be like trying to improve profits without producing financial accounts. Whilst one obviously must use a scale that is technically valid, the choice of rating scale for customer satisfaction measurement should be based on its suitability for achieving these two objectives rather than its validity for many other kinds of market 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’. Interval scales use numbers to distinguish the points on the scale. They are suitable for most statistical techniques 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.

Statisticians refer to two types of data – parametric and non-parametric statistics. In parametric statistics the data is a measured quantity, such as a volume of liquid, the speed of a vehicle, the temperature or, in research, numerical scores generated by interval or ratio scales. With parametric data you can draw bell curves (see Chapter 5), and a normal distribution is defined by two parameters, the mean and the standard deviation. Parametric, normally distributed data permits researchers to draw inferences about the extent to which a result from a sample applies to the whole population and can be analysed using multivariate statistics such as analysis of variance and covariance, regression analysis and factor analysis.

In non-parametric statistics, the data is not a measurable quantity of something but a count or a ranking, such as how many people have fair hair or black hair, how many times it was sunny or rainy or how many customers were satisfied or dissatisfied. Most types of non-parametric data are literally just counts of how many times something occurred, such as the number of days in the year that there was zero rainfall. Non-parametric statistics are analysed counting up the number of incidences, such as how many customers ticked the ‘satisfied’ or ‘very satisfied’ boxes, and this is known as a frequency distribution. From non-parametric statistics you can draw inferences about how many days it didn’t rain, or how many customers are satisfied, but not about the average rainfall or about the average level of satisfaction that your organisation is delivering. The statistical techniques used to analyse numbers can’t be employed with non-parametric statistics, which have to be analysed using counts and frequency distributions and reliability tests such as chi-square.
According to Allen and Rao:

“The use of ordinal scales in customer satisfaction measurement should be discouraged. It is meaningless to calculate any of the fundamental distributional metrics so familiar to customer satisfaction researchers. The average and standard deviation, for example, are highly suspect. Similarly, most multivariate statistical methods make assumptions that preclude the use of data measured on an ordinal scale.”

Without mean scores for importance and satisfaction it is not possible to calculate ‘satisfaction gaps’ – a huge handicap for the actionability of customer satisfaction surveys, nor a weighted index, which is the most accurate type of headline measure for judging success. Parametric statistics such as correlation or regression analysis are also used to establish the relationships between variables in the data set – an essential pre-requisite for understanding things like the drivers of satisfaction and loyalty. 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

Some people 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. The same people would often say that, by contrast, 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. It is true that the word ‘satisfied’ has more meaning than a score of 8/10. However, whilst people will ascribe a meaning to ‘satisfied’ that they are happy with, the problem is that it doesn’t have the same meaning. It is certainly undeniable that numbers, such as a score of 8/10 don’t have a meaning that people can readily interpret into words, but the fact that numbers do not have a meaning is their big strength for measuring. It’s why numbers were invented. Imagine if the early traders had to use a verbal scale such as ‘fair’, ‘very fair’ or ‘satisfied’, ‘very satisfied’ to judge the amount of wheat that should be traded for a horse. Numbers enable people to understand measures because they know not only that 4 pounds is better than 3 pounds, but how much better.

The same logic applies to satisfaction measurement. If a random sample of customers give a set of numbers this year to indicate their level of satisfaction and next year, another random sample gives a slightly higher set of numbers, we know they must be more satisfied. And provided the sample is large enough we can be sure that the higher level of satisfaction applies to the whole customer base. We may not be able to give a verbal meaning to our satisfaction index of 83.4% or our score of 7.65 for ‘ease of doing business’, but they will be truly accurate and comparable measures of the organisation’s success in delivering satisfaction from one year to the next. Moreover, since people interpret words in a variety of ways, it would be pointless to attempt to apply a verbal description to the scores achieved.

There are two particular types of CSM survey where numerical scales are much better than verbal scales for ease of completion. Firstly, numerical scales work much better for interviewing since respondents simply have to focus on giving a score, out of 10 for example, rather than struggling to remember the different points on the verbal scale, and anybody who has ever tried to interview customers using a verbal scale will know exactly how difficult it is, especially on the telephone. Secondly, for international surveys 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) The mid-point

Amateur researchers often worry about mid-points on rating scales, often believing that the mere existence of a mid-point encourages everyone to use it as an easy option. Evidence from CSM research totally contradicts this popular myth. The main difficulty of measuring importance is respondents’ tendency to score the higher points on the scale. As far as satisfaction is concerned, it is well known that customer satisfaction data is almost always positively skewed (see Section 6). However, the evidence from thousands of customer satisfaction surveys conducted by The Leadership Factor is that respondents score it how they see it, with relatively few going for the middle of the scale.
In theory one should include a mid-point on the scale, since it’s poor research to force respondents to express an opinion they don’t hold – they may genuinely be neither satisfied nor dissatisfied. However, we would have few concerns whether a scale had a mid-point or not. If you feel happier with 4 or 6 points rather than 5 or 7, we don’t believe it will make much difference, though bear in mind that since people don’t target the mid-point, there won’t be any detriment to including it either. Interestingly, a 10-point scale doesn’t have a mid-point, although this is academic since tests show that above 7 points, respondents typically don’t focus on where the mid-point is.

4) 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) or ‘top box’ scores (those ticking the highest point on the scale). As shown in Figure 1, 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 as a percentage satisfied, there is little point having more than 2 points on the scale – satisfied and dissatisfied.

5) Expectation scales

Some organisations use expectation scales in an attempt to measure the extent to which customers requirements have been met. Whilst these scales have some intuitive appeal, they suffer from three serious drawbacks for CSM. The first is that, like any verbal-type scale they have only ordinal properties so suffer from all the analytical limitations outlined above. A much bigger drawback, however, is their unsuitability as a benchmark for judging the organisation’s success. As pointed out by Grapentine⁴, if the measure changes in future is it because the company’s performance has improved or deteriorated or is it down to changes in customers’ expectations? In the same article, Grapentine also highlights the third problem with expectation scales for measuring customer satisfaction. For many ‘givens’, such as ‘cleanliness of the restaurant’, ‘accuracy of billing’ or ‘safety of the aeroplane’, customers never score above the mid-point. Whatever the level of investment or effort required to achieve them, clean restaurants, bills without mistakes and aeroplanes that don’t crash will never do more than meet the customer’s expectations.

6) Number of points

Since it is not practical to have many points on a verbal scale, 5-point verbal scales, are the norm. This is a considerable disadvantage since the differences between satisfaction survey results from one period to the next will often be very small.

As already mentioned in Section 3, one of the characteristics of CSM data is that it tends to be skewed towards the high end of the scale. This merely reflects the fact that companies generally perform well enough to make most customers broadly satisfied rather than dissatisfied – they wouldn’t still be there if they weren’t. (It is interesting to note that scores from situations where high levels of dissatisfaction do exist, typically when customers have no choice, do exhibit a much
more normal distribution). What most companies are mainly measuring therefore is degrees of satisfaction and since
they are 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 – and this is the big problem with five point
scales.

The problem is exacerbated by a tendency amongst many people to avoid the extremes of the scale. Even if we’re
mainly measuring degrees of satisfaction, this isn’t a major problem on a 10-point scale because there are still 4 options
(6, 7, 8 and 9) for the respondent who is reluctant to score the top box. With at least four choices it is therefore quite
feasible for customers to use the 10-point scale to acknowledge relatively small changes in a supplier’s performance. (As
an aside, when ‘top box’ scores on a 10-point scale are monitored it is normal practice to use 9s and 10s as ‘top box’
scores.) By contrast, it’s a big problem on a 5-point scale because for anyone reluctant to use the extremes of a scale
there’s only one place for the satisfied customer to go – and because so many people go there it has become known as
the ‘courtesy 4’! This often results in a narrow distribution of data with insufficient discrimination to monitor fine changes
in a supplier’s performance, so the slow, small improvements in satisfaction that one normally sees in the real world will
often be undetected by CSM surveys using 5-point scales. Consequently, whilst one can debate the rights and wrongs
of different scales from a pure research point of view, the disadvantages from a practical business management perspective
are obvious. Often it will lead to disillusionment amongst staff with the customer satisfaction process on the grounds
that “whatever we do it makes no difference, so it’s pointless trying to improve customer satisfaction”. It is therefore
essential from a business perspective to have a CSM methodology that is discriminating enough to detect any changes
in customer satisfaction, however small. As well as being more suitable for tracking small changes over time, scales
with more points discriminate better between top and poor performers so tend to have greater utility for management
decision making in situations where a company has multiple stores or branches or business units.

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.

![Figure 1: Distribution of data across scales](image-url)

It should also be noted that variance can be further improved on numerical scales by increasing the bi-polarity of the
anchored end points. A 10 point scale with end points labelled ‘dissatisfied’ and ‘satisfied’ would generate a narrower
distribution than end points labelled ‘very dissatisfied’ and ‘very satisfied’. Even better would be end points labelled
‘completely dissatisfied’ and ‘completely satisfied’.
Even from a technical research point of view there is a compelling argument for the 10-point scale because 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 customer loyalty, or establishing the relationship between employee satisfaction and customer satisfaction. In fact, many sophisticated statistical modelling packages assume that data are only ordinal if scales have fewer than 6 points.

In the light of 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 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 or from the reviews of footballers in newspapers. Numerical scales with fewer or more than 10 are more difficult for people as are verbal scales with more than 5 points.

Following a test at Cornell University in 1994, Wittink and Bayer concluded that the 10-point endpoint anchored numerical scale was most suitable for customer satisfaction measurement. Their reasons included respondent perspective issues such as simplicity and understandability as well as reliability issues such as repeatability (the extent to which the same scores are given by respondents in successive tests). Most importantly they concluded that it was the best scale for detecting changes over time and for improving customer satisfaction.

Michael Johnson from Michigan University Business School and Anders Gustafson form the Service Research Centre at Karlstad University (the originators of the American Customer Satisfaction Index and Swedish Customer Satisfaction Barometer) are in no doubt that a 10-point scale numerical scale should be used for customer satisfaction measurement.

7) The danger of over-stating customer satisfaction

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

According to Allen and Rao, a company that routinely scores 90% for overall customer satisfaction on a 5-point scale will typically score 85% on a 7-point scale and only 75% on a 10-point scale. This is corroborated by our own experience when we have 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 10-point numerical scale. This is done by duplicating some questions using both scales in the same questionnaire or in separate surveys conducted simultaneously. Sample data in Figure 2 from 5 example tests shows how results from the two different scales compared:

<table>
<thead>
<tr>
<th>Example 1</th>
<th>Example 2</th>
<th>Example 3</th>
<th>Example 4</th>
<th>Example 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>% “satisfied” (5 point verbal scale)</td>
<td>78.6%</td>
<td>84.5%</td>
<td>87.6%</td>
<td>90.3%</td>
</tr>
<tr>
<td>Satisfaction Index™ (10 point numerical scale)</td>
<td>65.3%</td>
<td>67.4%</td>
<td>70.9%</td>
<td>74.4%</td>
</tr>
</tbody>
</table>

*Figure 2: Satisfaction levels across scales*
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 very 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 misleading CSM process. Their customers are actually well below the levels of satisfaction that would generate loyalty.

**Over-stating customer satisfaction at AT&T**

The risk of over-estimating customer satisfaction by monitoring the top two boxes on 5-point scales was demonstrated by AT&T who 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). Using a weighted customer satisfaction index based on a 10-point scale, an Index of 90% would place you well into the top 5% of organisations on customer satisfaction, so if you’re regularly getting well over 90% you should smell a rat and, like AT&T question the validity of your CSM methodology.

A huge amount of evidence collected over a 30 year period by Harvard Business School also supports this view. 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.

---

**Figure 3: Satisfaction-Loyalty relationship**

---

[Diagram showing the satisfaction-loyalty relationship with zones of affection, indifference, and defection.]
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 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.

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.

Top box satisfaction

When General Motors, who were one of the pioneers of customer satisfaction research, began to suspect that their CSM process, (based on a balanced 5-point verbal scale) was not providing a sound basis for decision making they analysed 10 years of back data involving over 100,000 customer responses. It demonstrated that the relationship between customer satisfaction and loyalty was not linear but displayed the characteristics shown in Figure 3, with loyalty declining very rapidly if satisfaction scored anything lower than ‘very satisfied’. This led them to draw two conclusions. First, that only a ‘top box’ score represented an adequate level of business performance and second that having only one point on the scale that covered the whole range of performance from adequate upwards was clearly useless. Their solution to the latter problem was to invent a positively biased scale, to provide the basis for moving customers from satisfied to delighted:

- Delighted: ‘I received everything I expected and more’
- Totally satisfied: ‘Everything lived up to my expectations’
- Very satisfied: ‘Almost everything lived up to my expectations’
- Satisfied: ‘Most things lived up to my expectations’
- Not satisfied: ‘My expectations were not met’

Although the four levels of satisfaction are much more useful than the two offered by a balanced verbal scale, this is matched by the 10-point scale, which also has all the other analytical advantages of numerical over verbal scales. The top box concept, however, is interesting especially for high performing companies seeking to move ‘from good to great’. On a 10-point numerical scale it is generally considered that 9 is the loyalty threshold, so monitoring ‘top box’ scores (i.e. 9s and 10s) as well as an overall customer satisfaction index can be very useful.

9) National measures of 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 has been conducted through telephone interviews since 1994 and now covers 60% of the US economy, with quarterly scores for large companies, sectors and an ‘American Customer Satisfaction Index’ reflecting how satisfied Americans are with the totality of goods and services supplied by the American economy. Michigan Business School has conducted extensive modelling that links higher levels of customer satisfaction with better corporate financial performance and
produces a regular analysis of what’s happening to customer satisfaction across the USA and why it is going up or down. The questions on the ACSI are rated on a 10-point numerical scale. According to Michigan this for two main reasons. First because 10 points are needed to provide the required level of discrimination at the satisfied end of the scale (and a ten point verbal scale is not workable, especially for telephone interviews) and secondly because of the analytical benefits afforded by numerical scales.

When TLF was asked by the Institute of Customer Service in 2007 to recommend a methodology for the UKCSI, a 10-point numerical scale was recommended and adopted. UK businesses that want to benchmark themselves against the best companies in the UK would therefore be best served by using a 10-point numerical scale.

Conclusions

1) 5-point verbal and 10-point numerical scales are both easy for customers to complete.

2) Verbal and numerical scales differ massively in terms of the types of statistical techniques that are permissible, verbal scales possessing very limited analytical power.

3) Scales with more points generate more variance so are better for tracking satisfaction over time, for discriminating between high and low performers and for using sophisticated statistical techniques.

4) Scales with fewer points produce higher satisfaction results leading to complacency with the organisation and misunderstanding about why apparently ‘satisfied’ customers are defecting.

5) For organisations that want to give themselves the best chance of improving customer satisfaction as well as being able to reliably judge their success in achieving that goal, the 10-point numerical scale is the only sensible option for customer satisfaction measurement.

6) For that reason, it is the scale used by the University of Michigan for the American Customer Satisfaction Index and by the Institute of Customer Service for the UKCSI.

References: