Improving the Discrimination of SERVQUAL by using Magnitude Scaling

M.C.Hart
Department of Public Policy and Managerial Studies
De Montfort University
Leicester
LE7 9SU

Tel: *+44 (0)116   257   7780*
Fax: *+44 (0)116   257   7795*
email: *mchart@dmu.ac.uk*

World Congress - TQM In Action!   (Sheffield Hallam University, July 1996)

INTRODUCTION

The **SERVQUAL** survey instrument is one of the most widely utilised techniques for deriving a measure of the quality of service industries (Parasuraman, Zeithaml and Berry, 1985,1988; Zeithaml, Parasuraman and Berry, 1990). It has been extensively used, principally in the USA, for obtaining quantitative measures of consumer satisfaction. The instrument invites consumers to indicate the extent to which they agree with a series of statements which are designed to measure those elements of a service which consumers would expect as ideal (the Expectations score) and then those elements of a service that they have recently experienced (the Perceptions score) Satisfaction is then measured as the simple arithmetic 'gap' between Perceptions and Expectations (**S=P-E**). The scale is a composite of five dimensions (Tangibles, Reliability, Responsiveness, Assurance and Empathy). These dimensions will differ in salience as between different industries. In addition, consumers may place a higher value on some of these dimensions rather than others - for example, empathy might be more important in a hospital clinic but reliability in a bank. Respondents are also asked to indicate the importance of the relative weighting of the dimensions by allocating them a series of points which total to 100. In this way, it is possible for individual respondents to give much more weight to certain facets of the services under consideration than to others. A **SERVQUAL** score for each dimension can then be computed for each respondent by averaging the 'gap' scores for the questions relating to that dimension. It is then possible to produce a total weighted score for each customer by applying the weights that

the customer attaches to each dimension.   Finally, scores are averaged over the numbers of people sampled in any particular investigation.   Typical results will tend to show a negative score as expectations of an ideal service tend to run ahead of evaluations of a service as actually experienced.

THE NATURE OF THE SCALE DEPLOYED IN **SERVQUAL**

As previously mentioned,   respondents are asked to measure their levels of agreement or disagreement by   circling   a number in the range 1-7 where 1 represents complete disagreement with the statement in question whilst a 7 represents complete agreement. Although the authors of **SERVQUAL** do not explicitly discuss the point, it is evident that a scale deployed is essentially a Likert-type scale, which is ordinal in nature. However, conventional statistical orthodoxy would hold that arithmetic operations such as addition or subtraction are not legitimate (Siegel and Castellan, 1988). In the context of the **SERVQUAL** scale, a relevant question would be to ask whether the 'gap' between, for example, point 6 and point 7 on the scale is to be regarded in the same light as the difference between, say, point 1 and point 2 of the scale.   If the **SERVQUAL** scale is regarded as an interval or ratio scale rather than an ordinal scale, then the 'gap' between the points in both of the instances mentioned above would be the same i.e. 1.   However, it is not self-evident when respondents fill in rating scales that they deploy in their minds an essentially linear ratio scale.   The authors of **SERVQUAL** have made the assumption that it not illegitimate to derive gap scores by a process of subtraction.   Indeed, such a practice is common   and   one commentator has observed that many researchers have treated rating scales as interval data and failed to observe the minimal requirements for interval level measurement (Foddy,1993, pp.169-170). It is possible, therefore, that a scale such as **SERVQUAL** needs further refinement by exploring the meanings that respondents attach to indicating a point on a scale.

THE CONCEPT OF MAGNITUDE SCALING

Magnitude scaling is a term popularised principally by Lodge (1981) to indicate the process by which different points on a scale can be said to represent non-uniform weights or quantities.   The process derives from the principles of psychophysics in which respondents are asked to assign magnitudes to such physical sensations as the intensity of sound, brightness of light, the heaviness of objects and so on (Goldstein, 1989).   Drawing upon the principles of classical psychophysics, Lodge and his collaborators would take samples of respondents and instruct them in simple psychophysical tasks such as estimating the length of various lines, given the value associated with a given baseline (typically given a value of 50). If respondents can reliably estimate that a line of twice the length of the given baseline should have a value of 100, whilst one half the length should be given a value of 25, then the investigators could 'train' their subjects to think psychometrically by offering a series of lines of different lengths and asking respondents to estimate their lengths as fractions, or multiples, of the given baseline.   The geometric mean of the responses, when plotted against actual line lengths, typically showed straight-line associations when displayed on a log-log graph.
    Once it could be demonstrated that respondents have a reasonable competence in numeric line estimation, Lodge argue that through a process of 'cross-modality' it is possible to assign

similar magnitudes to points on an adjectival scale. Thus respondents when faced with a list of words such as 'So-So' (representing neutrality) could assign magnitudes to such concepts such as 'Good', 'Very Good' and 'Excellent' (multiples of the baseline) or to 'Bad','Very Bad' and 'Atrocious' (fractions of the baseline). Lodge presents results (from a small sample of 48 respondents) indicating the following magnitudes associated with a commonly used 'adjectives' on a scale:

Table 1 Magnitude Weightings assigned to common adjectives - Lodge,1981 *(n=48)*

	Atrocious	Very Bad	Bad	So-So	Good	Very Good	Excellent
	4	12	15	(50)	107	141	233
Multiple/ fraction	0.1	0.2	0.3	0	2.1	2.8	4.7

The final row of this table (added by the present author) indicates the extent to which each column is a multiple, or a fraction, of the given reference figure for neutrality or 'So-So' of 50. This data confirms the suspicion previously expressed that respondents do not use a linear interval scale when making judgements concerning the magnitudes to be assigned to adjectival descriptors of the various points of a rating scale.

The instructions that Lodge gave to his subjects allowed them to think of any multiple (or of any fraction) that they wished when assigning numerical weightings to adjectives. It is evident that the geometric mean of 233 indicates that most respondents felt that 'Excellent' represented 4-5 times as much weight as the reference point of 50. However, given that the any numbers on the right-hand scale could be in excess of 50 whereas any number less than 50 was automatically confined to the range 1-50, an argument could be made that the absence of a constraint on the one hand but the presence of it on the other altered the behaviour of the respondents when assigning weighting numbers.

In order to replicate the Lodge results and also to overcome the potential problem identified above, the exercise was repeated with a larger sample but with one or two crucial differences. The reference figure was given as 100 (rather than 50) and respondents were asked to supply figures *greater than* 100 to represent the weightings that they would attach to adjectives with a similar exemplar (reactions to a TV programme). The following results were obtained:

Table 2 Replication and refinement of Lodge Magnitude Weightings *(n=85)*

	Atrocious	Very Bad	Bad	So-So	Good	Very Good	Excellent
	267	201	143	(100)	187	269	362
Multiple/ fraction	2.6	2.0	1.4	0	1.9	2.7	3.6

It can be seen immediately that there is a fair measure of agreement in the magnitudes to be assigned to the adjectives 'Good' and 'Very Good' (approx. 2.0 and 2.8 respectively) whilst the magnitude associated with the word 'Excellent' shows more divergence (4.7 compared with 3.6). However, when it comes to the weightings assigned to the left hand side of the scale, the discrepancies become more severe. At first glance, the Lodge data appears to show that respondents only assign 1/12 of the intensity of support to an adjective

such as 'Atrocious' (4/50) whereas 'Excellent' represents 4-5 times as much support. Suspecting a scale effect, the data displayed in Table 2 probably represents a more accurate magnitude scaling of these common adjectives. It is very interesting in this case to observe that respondents are more inclined to be generous in expressing 'positive' support, whilst they are evidently less so when assigning magnitudes on the left hand side of the scale.

REFINEMENT OF **SERVQUAL** RATING SCALE

Table 3   Application of Magnitude Weightings to a **SERVQUAL** scale

	Strongly Disagree						*Strongly Agree*
Point on scale	1	2	3	4	5	6	7
Score of each point	-2.6	-2.0	-1.4	0	1.9	2.7	3.6

The **SERVQUAL** scale, as originally formulated in Zeithaml. et. al. (1990) only has the adjectival descriptors of 'Strongly Disagree' and 'Strongly Agree'. The data on magnitude scaling shown in Tables 1 and 2 above both suggest that rating scales cannot be regarded as interval scales and the subtraction of any two points on the scale would not give the results suggested by the framers of **SERVQUAL**. For example a movement from point 6 to 7 is 'worth' (3.6-2.7=0.9) whilst a corresponding movement from point 2 to point 1 is 'worth' (2.6-2.0=0.6) only two thirds of this. For this reason, it is suggested that when **SERVQUAL** type scales are administered, then the concepts of magnitude scaling be deployed in order to arrive at a series of putative magnitudes that can be assigned to different points on the scale, thus legitimating the original conception that satisfaction be conceptualised as the 'gap' between Perceptions of the service as delivered and Expectations of an ideal service.

REFERENCES

Foddy,W. (1993) *Constructing Questions for Interviews and Questionnaires*, Cambridge University Press,Cambridge.
Goldstein,B. (1989) *Sensation and Perception*. Wadsworth, Belmont,Ca.
Lodge,M. (1981) *Magnitude Scaling*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-025, Sage Publications, Beverly Hills and London
Parasuraman,A., Zeithaml,V. and Berry.L. (1985) A Conceptual Model of Service Quality and
     its implications for future Research, *Journal of Marketing*. **49** (Fall):41-50.
Parasuraman,A,. Zeithaml,V. and Berry,L. (1988) **SERVQUAL**: A Multiple-Item Scale for Measuring Perceptions of Service Quality, Journal of Retailing, 64: 12-40.
Siegel,S. and Castellan,N. (1988) Nonparametric Statistics for the Behavioural Sciences McGraw-Hill, New York
Zeithaml,V.,Parasuraman,A. and Berry,L. (1990) *Delivering Service Quality*, Free Press, New York