GROUNDED BENCHMARKS FOR SERVICE QUALITY

Catherine Eddy, Byron Sharp, Narelle Page and John Dawes
University of South Australia

Abstract

Competitive brands within a product category seldom differ much in overall customer perceived service quality. We report descriptive statistics (providing grounded benchmarks) for service quality. While we show that there are differences in average service quality scores between industries, competitive brands within an industry show almost no difference in mean service quality scores. Likewise their variation in individual scores is typically identical. This pattern, which runs completely counter to that assumed by the service quality literature, was found to hold across six industries, 56 brands, more than 30 countries and is based on in excess of 89,000 customer interviews using a wide range of survey and measurement approaches.

Introduction – the need for grounded benchmarks

The vast research literature on service quality is almost entirely devoted to (a) measuring service quality (eg, psychometrics) - (Asubonteng, McCleary, and Swan 1996; Babakus and Boller 1992; Carman 1990; Cronin and Taylor 1992; Cronin and Taylor 1994; Parasuraman, Zeithaml, and Berry 1988; Parasuraman, Zeithaml, and Berry 1994; Smith 1995); (b) developing and testing theories to explain service quality perceptions (e.g. Parasuraman, Zeithaml, and Berry 1985); and (c) developing and testing theories concerning the causal impact of service quality on things like customer loyalty and company profits (Anderson and Sullivan 1993; Anderson, Fornell, and Lehmann 1994; Rust and Oliver 1994; Rust and Zahorik 1993; Rust, Zahorik, and Keiningham 1995).

This is accompanied by a management literature largely devoted to exhortations for companies to improve service quality and case studies purporting to show how individual companies have achieved vast improvements in service quality.

This considerable body of theoretical work has been completed largely in a vacuum of knowledge concerning what service quality is actually like in the real world. We think that this is similar to early attempts to explain what 'stars' at night were (eg, angels, fireflies, celestial bodies or pinpricks of light from heaven itself shining through a ceiling) and how these stars related to one another (is there some system ?) without first documenting constellations and patterns in nightly, monthly and yearly movements across the sky.

By not documenting patterns in service quality a range of incredibly important assumptions have been implicitly made. For example, an assumption is that service quality varies considerably between competing brands (and hence is a possible explanation for differences in market share, customer retention or profitability). By implication there is the assumption that firms can and often do take action to drastically alter levels of service quality. Likewise that customers can change their expectations, presumably after receiving better service quality from another brand or in another product category. So it is assumed that service quality scores can and do change considerably over time, even without managerial action.

This seems to be an awful lot to simply assume. In short, the service quality literature is grounded on theoretical assumptions rather than empirical knowledge. Rather than seeking to explain real-world phenomena, the basic function of science, the emphasis seems to be on

building theories to explain an imaginary theoretical world, a world which may, or may not actually correspond to the real world that marketing takes place within.

Developing Grounded Benchmarks

Descriptive modelling, the classical task of science (Little 1994), tends to adopt quite a different approach than the usual modelling approaches employed in the marketing literature (see Ehrenberg, Barnard, and Sharp 2000 for discussion). The emphasis is not on fitting a 'best fit' model to a single set of data, but rather developing a model that fits across a wide range of conditions. So an initial task is to look across multiple data sets, covering various conditions, looking for invariant results.

We chose to look across a wide range of service quality surveys and to simply document the degree of variation in mean service quality scores obtained for competing brands. Our first question was "How much variation between brands is there ?" And, "How does it compare to the variation between product categories?"

This was not mindless empiricism, for we already know that loyalty rates vary little between competitive brands – a fact documented over several decades and numerous product categories (Ehrenberg 2000; Ehrenberg, Goodhardt, and Barwise 1990). And it is often held that service quality affects loyalty (Rust, Zahorik, and Keiningham 1995), so we wondered if service quality might really vary much between brands. We felt that the variation between product categories might give a good comparison point.

We found very little variation in either the mean scores obtained for competing brands or in the distribution of each brands' service quality scores. In comparison we found much greater variation between industries.

Our Data and Results

All of the data contained in our analysis resulted from commercially funded research projects. Approximately one third was collected in Australia and the remainder among 32 different overseas countries.

In most of the studies examined, an 11 point verbally anchored agreement scale was used to elicit an evaluation. The question was agreement with the statement "The service I receive from X is excellent" or alternatively in some data sets, it was the extent of satisfaction, again rated on an 11 point zero to ten scale (we found such high correlation between service quality and satisfaction scores within studies that we decided to use these measures interchangeably). In the data collected for hotels, the scale used was a 6 point overall satisfaction evaluation ("please indicate your overall satisfaction with X"). In order to bring this data in line with the other industries, the hotel evaluations were converted to an 11 point scale. So, 0 remained 0, 1 became 2, 2 became 4, etc. This was a reasonable strategy given that the purpose of this research was to create benchmarks and uniformity was desirable. Also, the hotel data was *all* 6 point scale data so we were making comparisons within the industry on like scales. The fact that we did *not* use identical scales across all data sets strengthens the generalisability of our findings, as they do not appear to be artefacts of a particular measurement device.

The results for an individual industry (table 1) show some variation between overall mean score. The fast food category exhibits the lowest average scores, while hotels seem to attract higher evaluations. This may not be surprising to many, however, it is an interesting and useful finding in that it provides a meaningful benchmark or norm with which comparisons can be made with other data (Churchill 1979).

Table 1 Mean Scores by Industry

Industry	n	Mean /10	No of brands
Hotels	57361	8.5	5
Building products	3332	7.7	8
Utility	6520	7.2	5
Banks	6789	7.0	15
Telco	7744	6.6	3
Fast food	7640	5.3	20
	89386	*7.8*	*56*

The distribution of scores for each industry is displayed in Chart 1. The black line in the box plot represents the median score. The mean score is displayed on the X axis labels. Both median and mean scores display differences across industry. We then show scores for brands within three of the industries (space limitations mean we cannot show this for all of the industries). The following three charts show the distribution and median scores (and mean scores in brackets) for competing brands within an industry.

Chart 1 Scores vary a lot across industries …..

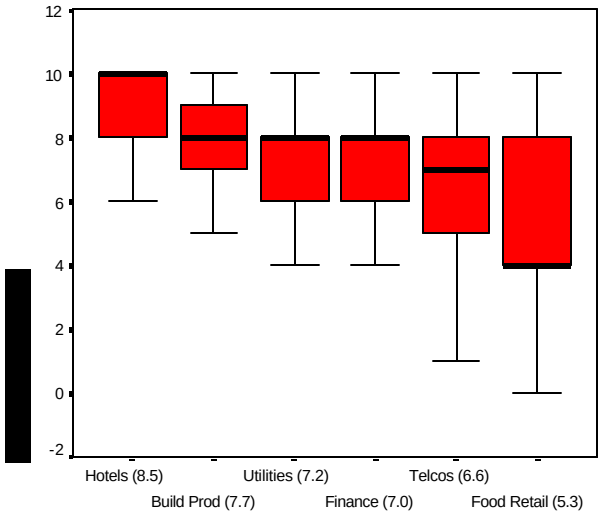


Chart 2 ….. but not by much within this industry

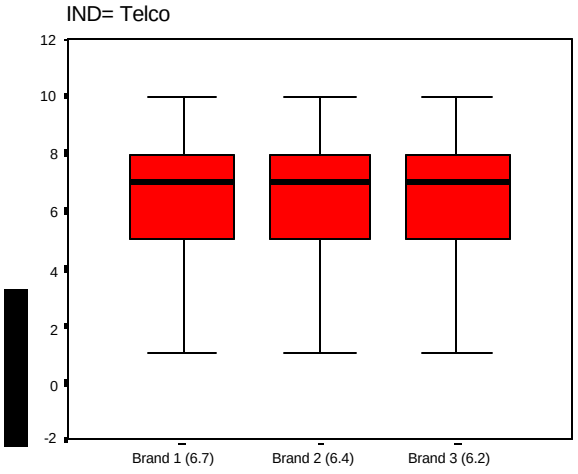


Chart 3 …. Little service variation between brands

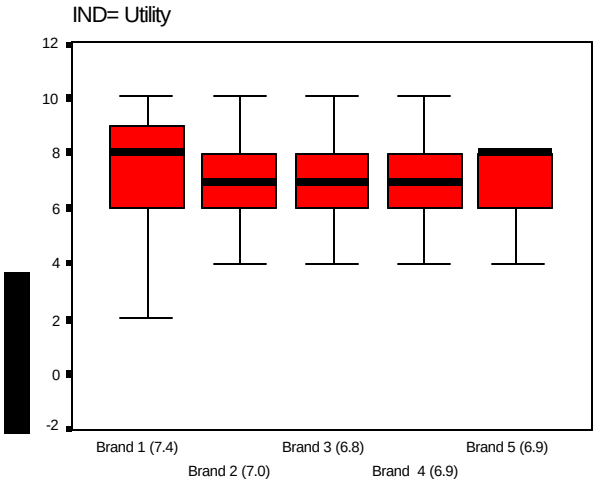
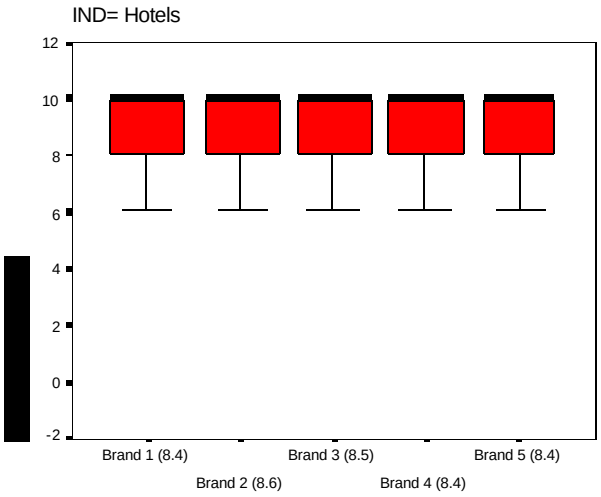


Chart 4 …. And the same again in this industry.



Discussion

The analysis shows that among competing brands in telecommunications, utilities and hotel industries there is almost no difference in mean or median score, or the distribution of those scores. We found similar results in the other industries - building products, finance, and fast food.

Conclusions / Implications

It is important to note that this finding does *not* suggest that service quality or the monitoring of service quality is not important. Presumably, service levels between brands are similar because the managers of those brands work very hard at competitive matching. However, it does suggest that managers need to be realistic about setting objectives for service quality improvement. The industry that one is in appears to be a prime delimiter as to what sorts of service ratings a particular brand will get. This then limits benchmarking against "world's best practice" if the firm is in another industry.

Since it appears that service scores should *not* be expected to differ by much, when research findings do show up sizeable differences, this should be grounds for thinking *something is wrong*. Either there has been a temporary service failure or a mistake in the data collection.

Finally, managers should be concerned about the degree of downward variation in service evaluations. We see in Chart 3 that brand 1 has a greater number of scores below the mid point of the scale (ie some scores of 4 out of 10 or less). This takes the level of analysis beyond the aggregate to the micro view. It may be, for instance, that these customers who perceive poor service are ones that are very important to the firm! It would also be reasonable to suggest that those customers providing very low service / satisfaction scores are more likely to defect to another firm. This in turn provides a possible explanation for why these ratings are similar across brands. Customers who are very dissatisfied can / do change to another firm, so this keeps the pattern of scores reasonably similar across the brands.

In conclusion, grounded benchmarks allow us to determine what is expected in service quality measurement and thereby provides a framework to interpret results. This work also allows managers to develop realistic service quality program objectives.

References

Anderson, Eugene and Mary W. Sullivan (1993), "The Antecedents and Consequences of Customer Satisfaction for Firms," *Marketing Science*, 12 (No. 2, Spring), 125-43.

Anderson, Eugene W., Claes Fornell, and Donald R. Lehmann (1994), "Customer Satisfaction, Market Share and Profitability: Findings from Sweden," *Journal of Marketing*, 58 (July), 53-66.

Asubonteng, Patrick, Karl J. McCleary, and John E. Swan (1996), "SERVQUAL Revisited: A Critical Review of Service Quality," *The Journal of Services Marketing*, 10 (No. 6), 62-81.

Babakus, Emin and Gregory W. Boller (1992), "An Empirical Assessment of the SERVQUAL Scale," *Journal of Business Research*, 24, 253-68.

Carman, J.M. (1990), "Consumer Perceptions of Service Quality: An Assessment of the SERVQUAL Dimensions," *Journal of Retailing*, 66 (Spring), 33-35.

Churchill, Gilbert A., Jr (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (February), 64-73.

Cronin, J. Joseph, Jr. and Steven A. Taylor (1992), "Measuring Service Quality: A Reexamination and Extension," *Journal of Marketing*, 56 (July), 55-66.

---- (1994), "SERVPERF Versus SERVQUAL: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality," *Journal of Marketing*, 58 (January), 125-31.

Ehrenberg, A.S.C., N.R. Barnard, and B. Sharp (2000), "Decision Models or Descriptive Models?," *International Journal of Research in Marketing*, 17 (2-3, September), 147-58.

Ehrenberg, Andrew S.C., Gerald J. Goodhardt, and T. Patrick Barwise (1990), "Double Jeopardy Revisited," *Journal of Marketing*, 54 (July), 82-91.

Ehrenberg, Andrew S.C. (2000), "Repeat Buying - Facts, Theory and Applications," *Journal of Empirical Generalisations in Marketing Science*, 5, 392-770.

Little, John D.C. (1994), "Modeling Market Response in Large Customer Panels," in *The Marketing Information Revolution*, Robert C. Blattberg and Rashi Glazer and John D.C. Little, Eds. Boston, Massachusetts: Harvard Business School Press.

Parasuraman, A, V Zeithaml, and LL Berry (1988), "SERVQUAL: A Multiple-item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, 64 (Spring), 12-40.

Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry (1985), "A Conceptual Model of Service Quality and Its Implications for Future Research," *Journal of Marketing*, 49 (Fall), 41-50.

---- (1994), "Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for Further Research," *Journal of Marketing*, 58 (January), 111-24.

Rust, Roland and Richard Oliver Eds. (1994), *Service Quality: New Directions in Theory and Practice*. London: Sage.

Rust, Roland T. and Anthony J. Zahorik (1993), "Customer Satisfaction, Customer Retention, and Market Share," *Journal of Retailing*, 69 (No. 2, Summer), 193-215.

Rust, Roland T., Anthony J. Zahorik, and Timothy L. Keiningham (1995), "Return on Quality (ROQ): Making Service Quality Financially Accountable," *Journal of Marketing*, 59 (April), 58-70.

Smith, Anne M. (1995), "Measuring Service Quality: Is SERVQUAL now Redundant?," *Journal of Marketing Management*, 11 (Jan/Feb/April, 1-3), 257-76.