

## Research Note:

# More on Improving Service Quality Measurement

**A. PARASURAMAN**

Federated Professor of Marketing  
College of Business Administration  
Texas A&M University  
College Station, Texas

**LEONARD L. BERRY**

J. C. Penney Chair of Retailing Studies  
College of Business Administration  
Texas A&M University  
College Station, Texas

**VALARIE A. ZEITHAML**

Partner  
Schmalensee/Zeithaml  
Chapel Hill, North Carolina

*In this note we respond to Brown, Churchill, and Peter's (1993) critique of SERVQUAL's difference-score conceptualization. We demonstrate that the claimed psychometric superiority of the alternative non-difference score conceptualization is debatable. We also argue that the SERVQUAL conceptualization offers richer diagnostics.*

Brown, Churchill, and Peter (1993) have written a thoughtful critique of SERVQUAL, an instrument for measuring service quality that we developed in 1988 (Parasuraman, Zeithaml, and Berry 1988) and later refined (Parasuraman, Berry, and Zeithaml 1991).<sup>1</sup> The primary focus of BCP's critique is the difference-score (i.e., perception minus expectation) conceptualization invoked by SERVQUAL to operationalize service quality. Specifically, they voice psychometric concerns about this conceptualization and, based on an empirical evaluation of SERVQUAL and an alternative non-difference score measure, conclude that the latter is superior. In this note we respond to their critique by addressing their concerns, raising

questions about several of their interpretations, and introducing additional issues that must be considered in comparing alternative scale formats. We argue that the alleged psychometric deficiencies of the difference-score formulation are not as severe as BCP suggest, and that the richer diagnostics of SERVQUAL may more than justify the separate measurement of perceptions and expectations. Our intent is to present additional perspective on the issues involved in choosing the most appropriate approach for assessing service quality.

## PSYCHOMETRIC ISSUES

### Reliability

As BCP correctly point out, the reliability of a scale operationalized as the difference between two measures will be low *to the extent that*: (a) the correlation between the component measures is high and/or (b) the reliabilities of the component measures are low. However, as discussed below, the conceptual definition of service quality (PZB 1985, 1988) as well as findings from empirical research (including our own as well as BCP's) suggest that conditions (a) and (b) above are not likely to be serious threats *when the construct being operationalized is an expectation-minus-perception difference score*.

The expectations component of SERVQUAL is a general measure and pertains to customers' *normative* standards—i.e., the service levels customers believe excellent companies in a sector must deliver. The perceptions component, on the other hand, pertains to customers' perceptions of a given company's service within the sector. As such, there is no conceptual reason for a customer's general evaluation standards to be correlated with his or her company-specific assessments. For example, if Customer A has a higher SERVQUAL expectation score for the appearance of department stores than does Customer B, it does not necessarily follow that Customer A would also rate XYZ department store's appearance higher than would Customer B. Any observed correlation between the SERVQUAL expectations and perceptions may be merely an artifact of both measures appearing on the same instrument (i.e., shared method variance). Such a correlation is not likely to be high, as evidenced by the moderate value of .34 in BCP's study (the correlations in our studies are also of similar magnitude).

Moreover, contrary to condition (b) above (i.e., low reliabilities of the component measures), BCP's study showed very strong reliabilities for the two components of SERVQUAL (.94 for expectations and .96 for percep-

tions). Our own findings from multiple studies in multiple sectors have demonstrated high reliabilities for the SERVQUAL measures.

### **Discriminant Validity**

BCP discuss two potential problems pertaining to the discriminant validity of difference-score measures. The first problem is that a difference-score measure's discriminant validity may be inflated *if the measure has low reliability*. Because the reliability of the SERVQUAL formulation has been shown to be consistently high (e.g., .87 to .92 in PZB 1988 and .94 in BCP 1993), this problem is unlikely to surface in studies using the difference-score formulation of SERVQUAL.

The second problem in BCP's view is that a difference-score measure would necessarily lack discriminant validity because it will be correlated with its two components. Although it is true that any difference-score measure is likely to be correlated with its components, we disagree with the inference that such a correlation demonstrates lack of discriminant validity, especially for the difference-score formulation of SERVQUAL. According to BCP's own definition, "discriminant validity refers to the degree to which measures of *theoretically unrelated constructs* do not correlate highly with one another" (emphasis added). Nowhere in our conceptualization (PZB 1985) or operationalization (PZB 1988) of service quality do we imply that the service quality construct is theoretically unrelated to expectations and perceptions. In fact, we state in PZB (1985) that service quality is a *function of* the discrepancy between customers' expectations and perceptions, implying that the former construct is necessarily related to the latter two. Thus inferring poor discriminant validity for the difference-score formulation of SERVQUAL on the basis of its correlation with its components is inconsistent with the definition of discriminant validity and hence inappropriate.

More appropriate criteria for assessing the discriminant validity of SERVQUAL are the correlations of its difference-score and non-difference score formulations with measures of the theoretically unrelated construct of consumer discontent. As BCP point out, both SERVQUAL formulations demonstrate discriminant validity on these criteria. In fact, it is noteworthy that the difference-score formulation displays somewhat stronger discriminant validity than the non-difference score formulation—the average magnitude of the correlations with consumer discontent is .04 for the former and .06 for the latter. While the difference between the two average correlations is small, it is similar to the differences on several other criteria

(e.g., reliability, nomological validity) that BCP invoke to claim psychometric superiority for the non-difference score formulation.

### **Variance Restriction**

BCP raise a legitimate concern that the high mean value and low standard deviation for the expectations component of SERVQUAL relative to the perceptions component will restrict the variance of the difference scores at higher levels of service quality. However, the relevance and seriousness of this potential problem depend on how the difference scores are used. The problem is mainly an issue when the difference scores are used in multivariate analyses. It is not relevant when the difference scores are used for diagnostic purposes: to pinpoint the most serious shortfalls along the general SERVQUAL dimensions or specific service attributes. Indeed, this type of diagnostic application of SERVQUAL dominates commercial use of the instrument and is one of its primary advantages (PBZ 1991).

Even when SERVQUAL scores are used in multivariate applications such as regression analysis, the variance restriction problem is likely to be serious *only if* the difference-score measure is the dependent variable, a point implied but not emphasized by BCP. More importantly, as BCP suggest, techniques such as generalized least-squares regression and variable transformations can be used to overcome this problem when it occurs. A possible reason why “none of the previous studies employing SERVQUAL used [these techniques]” as BCP indicate is that the difference scores served as independent variables in the regression analyses, if any, conducted in those studies (e.g., this was the case in all studies we cite and discuss in PZB 1991).

### **Other Psychometric Issues**

In assessing convergent validity based on the correlations in Table 1, BCP simply state that “the SERVQUAL measure, the non-difference score measure, and both global measures of service quality all correlate .60 or better, suggesting that all are measuring the same construct.” This statement, while true, fails to acknowledge that the SERVQUAL measure actually exhibits stronger convergent validity than the non-difference score measure as evidenced by its higher correlations with both the 1-item global measure (.67 vs. .63) and the 5-item global measure (.79 vs. .74). Though these differences are small they are of the same order of magnitude as other differences that BCP explicitly highlight in instances where the non-difference score measure appears to have an advantage.

The mean values of the SERVQUAL and non-difference score measures reported in Table 1 of BCP (1993) also warrant highlighting as they are germane to the comparative assessment of the two measures. The mean of  $-0.82$  for the SERVQUAL measure implies that on average respondents' perceptions fell short of their expectations, a logical finding given that the expectations component of service quality represents a form of "ideal" standard. In contrast, the mean of  $4.51$  for the non-difference score measure implies the opposite conclusion because this value falls between the "neutral" and "slightly better than expected" positions on the 7-point scale used. Since the same respondents rated their respective financial institutions using both measures, the most plausible explanation for these conflicting implications is that "expectations" as captured by BCP's non-difference score measure may be representing a lower standard than expectations as conceptualized in the service quality literature and operationalized in SERVQUAL. Thus there is some question about the face validity of BCP's non-difference score measure.

A related issue is the shape of the distributions of the two measures. As BCP correctly conclude, the non-difference score measure has a distribution that is much less skewed than the SERVQUAL measure's distribution. However, this finding also calls into question the face validity of the non-difference score measure: Respondents' perceptual ratings relative to expectations are more likely to follow a normal distribution when the expectation standard is low than when it is high. Therefore the choice between the two measures on the basis of their distributional properties is not as clear-cut as BCP's critique might imply.

The soundness of the five-dimensional framework originally proposed for the SERVQUAL items is another issue raised by BCP. We too have raised this issue and discussed it extensively in PBZ (1991) where we offer a comparative discussion of findings from our studies and those of other researchers who have evaluated SERVQUAL (Babakus and Boller 1992; Brensinger and Lambert 1990; Carman 1990; and Finn and Lamb 1991). Our comparative assessment acknowledges the mixed findings about SERVQUAL's factor structure, presents additional evidence and rationale supporting the viability of the five-dimensional framework, and provides directions for future research on this issue.

## **PRACTICAL ISSUES**

At the end of their critique, BCP wonder whether SERVQUAL has universal applicability, an issue we have addressed in PBZ (1991). As we

argue in that paper, the SERVQUAL items represent *core* evaluation criteria that transcend specific companies and industries. The SERVQUAL items are the basic “skeleton” underlying service quality that can be supplemented with context-specific items when necessary (guidelines for incorporating such items into the SERVQUAL framework are given in PBZ 1991).

Furthermore, BCP’s assertion that SERVQUAL suffers from the omission of critical items such as “the convenience of the bank’s location or its operating hours” is inaccurate. SERVQUAL *does* have an item focusing on the convenience of operating hours: “XYZ has operating hours convenient to all its customers” (PBZ 1991, item 19 in the Appendix).

Other practical issues in BCP’s critique relate to: (a) increased questionnaire length due to the separate expectations component and (b) the contribution of this component to explaining variance in other variables. While the first issue is a legitimate concern, our extensive experience with the two-part SERVQUAL instrument indicates that the benefits of richer, more accurate diagnostic information provided (as we demonstrate below) outweigh increased questionnaire length. Moreover, questionnaire length can be reduced by using just one list of the SERVQUAL items and placing the expectations and perceptions rating scales in two columns adjacent to the list. We are currently evaluating such a format in our ongoing efforts to refine and enrich the measurement of service quality.

Regarding the contribution of the expectations component to explaining variance in other variables, our findings are similar to those of other researchers—i.e., the perceptions measure seems to outperform the difference-score measure in explaining the variance in other variables. Ironically, there is evidence of exceptions to this pattern in BCP’s own study. As their Table 1 reveals, in terms of explaining the variance in the global service quality measures, SERVQUAL outperforms not only the non-difference score measure but also the perceptions measure (.67 to .63 for the 1-item measure and .79 to .77 for the 5-item measure).

The more important trade-off question to ask here is the following. Is the increased ability to explain variance worth the potential loss of richer, more accurate diagnostics for improving service quality? Findings from our studies (described in PBZ 1991) involving independent customer samples of five nationally-known companies strongly suggest that the perceptions ratings alone may not lead to the same (or correct) practical implications as the perceptions-expectations difference scores.

For instance, consider the following mean perceptions and SERVQUAL scores obtained by an insurance company for the five service quality dimensions:

<b>Dimension</b>	<b>Perceptions Scores</b>	<b>SERVQUAL Scores</b>
Tangibles	5.3	0.0
Reliability	4.8	- 1.6
Responsiveness	5.1	- 1.3
Assurance	5.4	- 1.0
Empathy	5.1	- 1.1

The perceptions ratings suggest placing equal emphasis on improving responsiveness and empathy when, in fact, the company has a bigger problem with responsiveness as the SERVQUAL scores reveal. This company would also focus more attention on improving its tangibles than on enhancing assurance if it relied solely on the perceptions scores. Clearly, this would be a major mistake as indicated by the SERVQUAL scores for tangibles and assurance.

Measuring expectations and perceptions separately also allows managers to better understand the dynamics of customers' assessments of service quality over time. For example, if SERVQUAL scores for certain items have declined significantly from one period to another, managers can assess whether this is due to higher expectations, lower perceptions, or both. This information is not available when perceptions relative to expectations are measured on the same scale.

Another advantage of measuring expectations and perceptions separately is that the gathered data can serve equally well the dual objectives of accurately diagnosing service shortfalls and explaining the variance in related variables. Difference scores can be used for the former while perceptions scores alone can be used for the latter (although, as already mentioned, the available empirical evidence does reveal exceptions to the presumed superiority of perceptions scores in explaining variance).

In the latest phase of our ongoing service-quality research we have measured two levels of expectations—adequate and desired. Managers using this approach can learn whether their customers' service perceptions fall within a "zone of tolerance" (the space between adequate service and desired service) or outside the zone (see ZBP 1993). Managers can see where customers' perceptions fall relative to the zone of tolerance for individual service quality items and dimensions, and compare their own customer data to competitor customer data. These insights are possible only if customers' expectations are measured separately.

## CONCLUSION

BCP's contribution to the literature on service-quality measurement is laudable. Their study is one of the first to assess empirically two alternative scale formats, *both of which incorporate customer expectations*, on an important set of psychometric criteria. However, as we hope we have demonstrated in our response and clarification of the issues, the collective conceptual and empirical evidence neither demonstrates clear superiority for the non-difference score format nor warrants abandoning the difference-score format as BCP's critique might imply.

## REFERENCES

- Babakus, Emin, and Gregory W. Boller (1992), "An Empirical Assessment of the SERVQUAL Scale," *Journal of Business Research*, **24**, 253–68.
- Brensinger, Ronald P., and Douglas M. Lambert (1990), "Can the SERVQUAL Scale be Generalized to Business-to-Business Services?" in *Knowledge Development in Marketing*, 1990 AMA's Summer Educators' Conference Proceedings, 289.
- Brown, Tom J., Gilbert A. Churchill, Jr., and J. Paul Peter (1993), "Improving the Measurement of Service Quality," *Journal of Retailing*, **69** (Spring).
- Carman, James M. (1990), "Consumer Perceptions of Service Quality: An Assessment of the SERVQUAL Dimensions," *Journal of Retailing*, **66** (Spring), 33–35.
- Finn, David W., and Charles W. Lamb, Jr. (1991), "An Evaluation of the SERVQUAL Scales in a Retail Setting," in Rebecca H. Holman and Michael R. Solomon (eds.), *Advances in Consumer Research*, Vol. 18, Provo, UT: Association for Consumer Research.
- Parasuraman, A., Leonard L. Berry, and Valarie A. Zeithaml (1991), "Refinement and Reassessment of the SERVQUAL Scale," *Journal of Retailing*, **67** (Winter), 420–50.
- , 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.
- , ———, and ——— (1988), "SERVQUAL: A Multiple-Item Scale for Measuring Customer Perceptions of Service Quality," *Journal of Retailing*, **64** (Spring), 12–40.
- Zeithaml, Valarie A., Leonard L. Berry, and A. Parasuraman (1993), "The Nature and Determinants of Customer Expectations of Service," *Journal of the Academy of Marketing Science*, **21** (Winter), 1–12.