

DIMENSIONALITY, RELIABILITY AND VALIDITY OF SERVQUAL[†]

Ugur Yavas ·
Zeynep Bilgin ··

This article examines selected psychometric properties of SERVQUAL in an international setting. An empirical study conducted among Turkish students reveals that SERVQUAL demonstrates acceptable reliability and validity but needs internal structure refinement, since different dimensions load on the same factors.

A deliberate attempt to study services marketing dates back to the mid-1960s (Rathmell, 1966). However, the interest on the topic has gained considerable momentum within the past decade, undoubtedly owing to Parasuraman, Zeithaml and Berry's (1985, 1988) seminal work on service quality. In today's competitive markets, businesses seek profitable ways to differentiate themselves and to gain a competitive edge over their rivals. Delivery of high service quality to customers offers firms an opportunity to distinguish themselves in crowded markets. Unlike goods quality, which can be measured objectively, service quality is abstract and elusive. In the absence of objective measures, firms must rely on consumers' perceptions of service quality to determine their own relative strengths and weaknesses, and to set up priorities. Hence, development of valid instruments to measure service quality is imperative.

Parasuraman, Zeithaml and Berry's (1985) initial work toward development of a service quality measure was based on in-depth interviews with executives and twelve focus groups with consumers in four service industries (credit card, banking, brokerage, and repair services). This

[†] This research was supported by a Non-Instructional Assignment and a Research Development Committee grant awarded to Dr. Yavas by East Tennessee State University.

* Ugur Yavas is a Professor of Marketing at East Tennessee State University, Johnson City, Tennessee.

** Zeynep Bilgin is an Associate professor of Marketing at Marmara University, Istanbul, Turkey.

Manuscript received, April, 1996, revised, July, 1996.

pioneering effort identified ten determinants of service quality including access, communication, competence, courtesy, credibility, reliability, responsiveness, security, tangibles and understanding/knowing customers. In subsequent work, by using an iterative procedure via factor analysis of a ninety-seven-item questionnaire, Parasuraman, Zeithaml and Berry (1988) refined the ten determinants into SERVQUAL, an instrument specifically designed to measure service quality.

In SERVQUAL, the initial ten determinants were consolidated into five: tangibles (TANG)--physical evidence of the service; Reliability (RELI)--consistency of performance and dependability; Responsiveness (RESP)--willingness and readiness of employees to provide service; Assurance (ASSU)--confidence communicated by the service provider, and Empathy (EMPA)--service provider's efforts to understand the customer's needs and then to individualize the service delivery. Assurance encompassed the prior five determinants of communication, credibility, security, competence and credibility, and EMPA captured the former access and knowing/understanding the consumer dimensions. Parasuraman, Zeithaml and Berry proposed that SERVQUAL was an adaptable instrument which could fit any organization's needs in measuring service quality. The considerable enthusiasm spawned by SERVQUAL is evidenced by numerous works which examined its psychometric properties and/or used it in applied settings (see, for example: Cronin and Taylor, 1992, 1994; Carman, 1990; Gagliano and Hathcote, 1994; Brown and Swartz, 1989).

The purpose of this article is to examine dimensionality, reliability and validity of the wide-ranging SERVQUAL instrument in the Turkish setting. Specifically, by using a sample of Turkish college students as its database, this study expands upon the research by Yavas and Arsan (1995) and Akan (1995) who examined psychometric properties of SERVQUAL among bank employees and consumers in Turkey.

METHOD

Sample

Data for the study were collected from undergraduate students attending Bogazici and Marmara universities located in Istanbul. Students in both schools completed the questionnaire in a self-administered manner during regular class periods. The medium of instruction at Bogazici University is English. Marmara University, in its separate divisions, provides instruction in Turkish and English. To eliminate the cross-linguistic equivalence problem associated with questionnaire translation (Aulakh and Kotabe, 1993), the surveys in the latter institution were administered only to the students attending the English division. Usable responses were obtained from 292 students. The sample was almost evenly divided in gender--

JOURNAL OF BUSINESS AND MANAGEMENT

female (51.7 percent) and male (48.3 percent). Students were primarily business/economics majors (61.3 percent).

Use of students as surrogates of other populations and generalizability of results obtained from student samples are often questioned (Burnett and Dunne, 1986). However, evidence shows that both in domestic and international studies, use of student subjects is appropriate when the objective is to assess psychometric properties of instruments (Dunne, Lund and Luchsinger, 1980; Yavas, 1994).

Measurement

The final form of SERVQUAL is comprised of 22 items (Parasuraman, Zeithaml and Berry, 1988). The breakdown of the items by dimension is as follows:

TANG	(4)
RELI	(5)
RESP	(4)
ASSU	(4)
EMPA	(5)

In this study, service quality was operationalized by using these 22 items. Specifically, for each item, the subjects were asked to evaluate the performances of their bank, doctor, hair stylist (barber) and post office on seven-point scales ranging from "Much worse than I expected" to "Much better than I expected." This specific scale was borrowed from Brown, Churchill and Peter (1993) whose research indicated that it is more efficient than the disconfirmation procedure used in the original SERVQUAL measure.

RESULTS

Dimensionality

To investigate the purported dimensionality of the SERVQUAL instrument (i.e., TANG, RELI, RESP, ASSU and EMPA), two sets of factor analyses were run. First, for each service setting, 22 SERVQUAL items were subjected to exploratory factor analysis. None of these analyses yielded the five-structure solution posited by Parasuraman, Zeithaml and Berry (1988). Instead, three factors with eigenvalues greater than one emerged in the cases of bank, barber and

post office settings and a two-factor solution was obtained in the doctor setting. In the next step, factor analyses were run for each setting by restricting the number of factors to five. In other words, the factor program was forced to generate five-factor solutions. Tables 1, 2, 3 and 4 present these results for the bank, doctor, barber and post office settings, respectively.

As can be seen from Table 1, nine items had their highest loadings on Factor 1 which accounted for 45.8 percent of the variance in the data. While four of the five EMPA items loaded on this factor, the factor also had items from the other domains. Four RELI items along with two items from the RESP domain were at the root of Factor 2. Two of the four ASSU items loaded on Factor 3 and three of the four TANG items loaded on Factor 4 (see Table 1). Of all the domains, TANG appears to be the only one which is distinct from the others. As shown in Tables 2, 3 and 4, four items comprising this domain load on distinct factors. Otherwise, there are no clear-cut patterns in the compositions of other factors. For instance, Factor 1 in Table 2 includes one RELI, three RESP, four ASSU and four EMPA items. Six items with highest loadings on Factor 2 (Table 3) belong to three different domains, RELI, RESP and ASSU. The same is true for the second factor of the post office setting where seven items comprising the factor come from these three domains (see Table 4).

The preceding discussion suggests that the dimensionality of SERVQUAL instrument is suspect. The instrument does not decompose into the purported five-structure solution. Even when the analysis is forced to conform to a five-factor structure, composition of the factors are different than those intended in the original formulation.

Table 1

Factor Analysis: Bank^{a,b}

Domain	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
TANG	1	-	-	3	-
RELI	1	4	-	-	-
RESP	2	2	-	-	-
ASSU	1	-	2	-	1
EMPA	4	-	-	-	1
Percent of Variance	45.8	7.7	6.5	4.4	3.9

^a Results are based on varimax rotated matrix

^b Numbers across each domain indicate the number of items from that particular domain which have their highest loadings on that factor

Table 2

Factor Analysis: Doctor^{a,b}

Domain	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
TANG	-	4	-	-	-
RELI	1	-	4	-	-
RESP	3	-	1	-	-
ASSU	4	-	-	-	-
EMPA	4	-	-	1	-
Percent of Variance	67.8	5.8	4.2	2.5	2.2

^a Results are based on varimax rotated matrix

^b Numbers across each domain indicate the number of items from that particular domain which have their highest loadings on that factor

Table 3

Factor Analysis: Barber^{a,b}

Domain	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
TANG	-	-	-	4	-
RELI	-	2	3	-	-
RESP	-	2	2	-	-
ASSU	2	2	-	-	-
EMPA	5	-	-	-	-
Percent of Variance	56.4	9.3	6.6	3.1	2.8

^a Results are based on varimax rotated matrix

^b Numbers across each domain indicate the number of items from that particular domain which have their highest loadings on that factor

Table 4

Factor Analysis: Post Office^{a,b}

Domain	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
TANG	-	-	4	-	-
RELI	1	4	-	-	-
RESP	2	2	-	-	-
ASSU	2	1	-	1	-
EMPA	4	-	-	-	1
Percent of Variance	51.8	7.9	6.6	4.3	3.4

^a Results are based on varimax rotated matrix

^b Numbers across each domain indicate the number of items from that particular domain which have their highest loadings on that factor

The adverse dimensionality finding, however, is not surprising in factor analysis of complex structures. Furthermore, this study is not unique in its failure to generate the neat five-factor structure posited by Parasuraman, Zeithaml and Berry (1988). Cronin and Taylor's (1992) confirmatory factor analysis of the SERVQUAL items resulted in a unidimensional solution. The five-factor structure they obtained from oblique factor rotation had a poor fit. Finn and Lamb (1991) and Spreng and Singh (1993) used confirmatory factor analysis and their five-factor solutions had poor fits. Brensinger and Lambert (1990) generated a five-factor solution but only four factors had eigenvalues exceeding one. Babakus and Boller's (1992) analysis of 22 SERVQUAL items produced a two-factor structure. In one of the earlier studies conducted in Turkey, Yavas and Arsan (1995) obtained a five-factor solution, yet the decomposition of the items into the five dimensions was different than the one purported in SERVQUAL. The number of factors extracted by Akan (1995) in her study of Turkish consumers ranged from seven to twelve.

Reliability

To assess the reliability of the five dimensions of SERVQUAL as well as the overall instrument itself across settings, coefficient alphas were computed. Coefficient alpha indicates reliability in terms of the internal consistency of items relating to a multi-item scale (Peterson, 1994). Reliability coefficients of .70 or higher are deemed acceptably high (Nunnally, 1978). The internal consistency reliability coefficients at the dimension level ranged between .75 and

JOURNAL OF BUSINESS AND MANAGEMENT

.91 (TANG), .86 and .92 (RELI), .83 and .93 (RESP), .80 and .93 (ASSU), and .86 and .94 (EMPA). Furthermore, the coefficient alphas for the overall instrument across all four settings were well above Nunnally's (1978) guidelines; they ranged between .94 (bank) and .97 (doctor).

To further examine the internal consistency reliability of the measures, inter-domain and domain-total correlations were compared to the Guilford and Fruchter (1973) guidelines. These guidelines suggest that inter-domain correlations should fall within .10 - .60 bounds and domain-total correlations should range between .30 and .80. An examination of the results revealed that of the fifteen pairwise correlations computed for each service setting, doctor had the most and bank the least number of correlation coefficients falling outside of the suggested range. Yet even in the case of bank, two-thirds of the coefficients were out-of-bounds. In this as well as the other three settings, coefficients falling outside of the guidelines exceeded the upper bound.

An inspection of the results at the domain level showed that of all five domains, tangibles most conformed to the Guilford-Fruchter (1973) guidelines. Of the twenty inter-domain and domain-total correlations computed for tangibles across all four settings, fourteen were within the guidelines. With all correlations falling within bounds, TANG's best performances were in the cases of bank and barber settings.

Validity

Convergent Validity. Confirmation of the existence of a construct indicated by correlations of independent measures of the construct provides evidence for convergent validity (Jaccard, Brinberg and Ackerman, 1986). To test for convergent validity, for each setting, a single-item direct service quality measure was included in the survey. To obtain this measure, respondents were asked to rate the overall quality of services they received from their banks, doctors, barbers/hair stylists and post-offices on a six point scale ranging from excellent to terrible. These single-items measures were then correlated with their multiple-item counterparts (i.e., the overall service quality scores obtained for each respondent by summing their answers across 22 SERVQUAL items). All the correlations were significant at .0001 level of significance and they ranged from a high of .87 (doctor) to a low of .64 (bank). These results attest to the convergent validity of the instrument.

Nomological Validity. When a construct of interest is related to other constructs assessing a different but conceptually related construct by an established body of theory or according to a priori expectations (Peter, 1981), confirmation of the relationship predicted by the theory/expectations is evidence of nomological validity. Past writings (O'Connor, Shewchuk and Bowers, 1991; Boulding et al., 1993; Bitner, 1990; Brown and Swartz, 1989) show that

service quality is significantly related to consumer satisfaction, recommending, continued patronage/repeat purchase, complaining and intentions to switch.

In this study, nomological validity of SERVQUAL was assessed by relating it to four behavioral measures. These were consumer satisfaction, intention to continue to do business/patronize the service provider, recommending it to friends and complaint behavior. Satisfaction was operationalized by a five-item measure where each item was measured on a 6-point scale ranging from "extremely satisfied" to "extremely dissatisfied." Intention to patronize was measured on a 5-point "very likely" to "very unlikely" scale. Again a 5-point scale ranging from "definitely would recommend" to "definitely would not recommend" was employed in measuring the recommendation behavior. Complaint behavior was measured in terms of frequency of complaints via a 5-point scale. The anchor points of this scale were "always" and "never."

It was hypothesized that higher levels of service quality sentiments would result in higher levels of consumer satisfaction, higher likelihoods of continuing to patronize the particular service provider, higher probabilities of recommending that provider to friends and lower frequencies of complaint. Because of the scoring system used, the signs of the coefficients were expected to be positive with respect to the first three hypotheses and negative for the fourth one.

The results reported in Table 5 lend support to these expectations. All the correlation coefficients were significant at .05 or better level of significance. Furthermore the coefficients had the expected signs. These results suggest that SERVQUAL demonstrates reasonable nomological validity.

CONCLUSIONS

The past decade or so was characterized by rapid internationalization of business. As a result, markets in many industries are becoming increasingly integrated worldwide. Such developments stimulate interest in international research in general, and methodological issues surrounding cross-national research in particular. Measurement equivalence is one of the methodological issues which is gaining increased attention. A common a priori assumption in international marketing research has been that measures developed in one culture will be universally applicable to other cultures. Typically, emic measures developed in the United States are transferred to other cultures without any modification. It appears that by virtue of precedence, cross-validity of measures has been endorsed. However, such assumptions and endorsements are now being challenged.

JOURNAL OF BUSINESS AND MANAGEMENT

This study examined dimensionality, reliability and validity of a wide-ranging measure developed in the United States (SERVQUAL), in the Turkish setting. The results of the study show that SERVQUAL instrument demonstrates acceptable reliability and validity when tested in a different culture. While this is encouraging, results also suggest that caution should be exercised when using the instrument. It appears that the internal structure of the scale is not crystallized. Items purporting to measure different dimensions tend to load on the same factor, and many items contribute little to the assessment of different dimensions. Hence, further refinement of the scale is in order.

Table 5

Nomological Validity of SERVQUAL Across Settings^a

Measure	Bank	Doctor	Barber	Post Office
Continue to Patronize	.43	.78	.68	.39
Satisfaction	.77	.91	.87	.81
Recommend to Friends	.65	.82	.69	.57
Complain	-.45	-.62	-.42	-.44

^a All the correlation coefficients reported in the table are significant at .05 or better level of significance

To achieve this, Churchill's (1979) paradigm for developing better measures of marketing constructs, recently illustrated by Webster (1993) in a service context, can be followed. Briefly, the procedure includes: a) specifying the domain of the construct, b) generating sample items that may tap the construct, c) collecting data on the measures, d) purifying the measure via coefficient alpha and factor analysis, e) collecting additional data to further assess reliability/validity and f) developing norms. These steps are essential to refine the SERVQUAL measure in the U.S. setting where the instrument suffers from dimensionality problem. However, the systematic procedure is imperative for the Turkish setting not only to address the dimensionality issue but also to determine the scalar and conceptual equivalence of the SERVQUAL measure.

REFERENCES

- Akan, P. (1995). *Dimensions of Service Quality: Expectations of Turkish Consumers from Services*. Istanbul: Bogazici University.
- Aulakh, P.S. & Kotabe, M. (1993). "An Assessment of Theoretical and Methodological Developments in International Marketing." *Journal of International Marketing*, 1(2), 5-28.
- Babakus, E. & Boller, G.W. (1992). "An Empirical Assessment of the SERVQUAL Scale." *Journal of Business Research*, 24(3), 253-268.
- Bitner, M.J. (1990). "Evaluating Service Encounters: The Effects of Physical Surroundings and Employee Responses." *Journal of Marketing*, 54(2), 69-82.
- Boulding, W., Kalra, A., Staelin, R. & Zeithaml, V.A. (1993). "A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions." *Journal of Marketing Research*, 30(1), 7-27.
- Brensinger, R. & Lambert, D.M. (1990). "Can the SERVQUAL Scale be Generalized to Business-to-Business Services?" In Bearden, W. et al. (Eds.), *Enhancing Knowledge Development in Marketing* (p. 289). Chicago: American Marketing Association.
- Brown, S.W. & Swartz, A. (1989). "A Gap Analysis of Professional Service Quality." *Journal of Marketing*, 53(2), 92-98.
- Brown, T.J., Churchill, G.A. & Peter, J.P. (1993). "Improving the Measurement of Service Quality." *Journal of Retailing*, 69(1), 127-139.
- Burnett, J.J. & Dunne, P.M. (1986). "An Appraisal of the Use of Student Subjects in Marketing Research." *Journal of Business Research*, 14(4), 329-343.
- Carman, J.M. (1990). "Consumer Perceptions of Service Quality: An Assessment of the SERVQUAL Dimensions." *Journal of Retailing*, 66(1), 33-55.
- Churchill, G.A. (1979). "A Paradigm for Developing Better Measures of Marketing Constructs." *Journal of Marketing Research*, 16(1), 64-73.
- Cronin, J.J. & Taylor, S.A. (1992). "Measuring Service Quality: A Reexamination and Extension." *Journal of Marketing*, 56(3), 55-68.
- Cronin, J.J. & Taylor, S.A. (1994). "SERVPERF versus SERVQUAL: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality." *Journal of Marketing*, 58(1), 125-131.
- Dunne, P., Lund, D. & Luchsinger, V. (1980). "The Use of Students as Surrogates: A Jeopardy Game?" In Bagozzi, R.P. et al. (Eds.). *Marketing in the 80's: Changes and Challenges* (Pp. 312-315). Chicago: American Marketing Association.
- Finn, D.W. & Lamb, C.W. (1991). "An Evaluation of the SERVQUAL Scale in a Retailing Setting." In Holman, R. and Solomon, M.R. (Eds.). *Advances in Consumer Research* (Pp. 483-490). Provo: Association for Consumer Research.
- Gagliano, K.B. & Hathcote, J. (1994). "Customer Expectations and Perceptions of Service Quality in Retail Apparel Specialty Stores." *Journal of Services Marketing*, 8(1), 60-69.

JOURNAL OF BUSINESS AND MANAGEMENT

- Guilford, J.P. & Fruchter, B. (1973). *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill Book Company.
- Jaccard, J., Brinberg, D. & Ackerman, L. (1986). "Assessing Attribute Importance: A Comparison of Six Methods." *Journal of Consumer Research*, 12(4), 463-468.
- Nunnally, J.C. (1978). *Psychometric Theory*. New York: McGraw-Hill.
- O'Connor, S.J., Shewchuk, R.M. & Bowers, M.R. (1991). "A Model of Service Quality Perceptions and Health Care Consumer Behavior." *Journal of Hospital Marketing*, 6(1), 69-92.
- Parasuraman, A., Zeithaml, V.A. & Berry, L.L. (1985). "A Conceptual Model of Service Quality and Its Implications for Future Research." *Journal of Marketing*, 49(4), 41-50.
- Parasuraman, A., Zeithaml, V.A. & Berry, L.L. (1988). "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality." *Journal of Retailing*, 64(1), 12-37.
- Peter, J.P. (1981). "Construct Validity: A Review of Basic Issues and Marketing Practices." *Journal of Marketing Research*, 18(2), 133-145.
- Peterson, R.A. (1994). "A Meta-analysis of Cronbach's Coefficient Alpha." *Journal of Consumer Research*, 21(2), 381-391.
- Rathmell, J.M. (1966). "What is Meant by Services?" *Journal of Marketing*, 30(4), 32-36.
- Spreng, R.A. & Singh, A.K. (1993). "An Empirical Assessment of the SERVQUAL Scale and the Relationship between Service Quality and Satisfaction." In Cravens, D.W. and Dickson, P. (Eds.). *Enhancing Knowledge Development in Marketing* (Pp. 1-6). Chicago: American Marketing Association.
- Webster, C. (1993). "Refinement of the Marketing Culture Scale and the Relationship between Marketing Culture and Profitability of a Service Firm." *Journal of Business Research*, 26(2), 111-131.
- Yavas, U. & Arsan, N. (1995). "An Assessment of Selected Psychometric Properties of a Service Quality Scale: An International Study (Bank Employees)." *Journal of Customer Service in Marketing and Management*, 1(4), 105-116.
- Yavas, U. (1994). "Research Note: Students as Subjects in Advertising and Marketing Research." *International Marketing Review*, 11(4), 35-43.