

# Measuring IS System Service Quality with SERVQUAL: Users' Perceptions of Relative Importance of the Five SERVPERF Dimensions

**Hollis Landrum**  
*U.S. Army Engineer Research  
and Development Center—  
Retired, Little Rock, AK, USA*

[hlandrum@att.net](mailto:hlandrum@att.net)

**Xiaoni Zhang**  
*Northern Kentucky University,  
Highland Heights, KY, USA*

[zhangx@nku.edu](mailto:zhangx@nku.edu)

**Victor Prybutok**  
*University of North Texas,  
Denton, TX, USA*

[prybutok@unt.edu](mailto:prybutok@unt.edu)

**Daniel Peak**  
*University of North Texas,  
Denton, TX, USA*

[peak@unt.edu](mailto:peak@unt.edu)

## Abstract

Using a library information system in an engineering research environment, this paper investigates the service quality perceptions of professional information system users. This paper also examines the relationship between the relative importance allocated by the system's users to each of the five SERVPERF dimensions (dimensions that are shared with the SERVQUAL instrument) as measured by the points allocated to that dimension and the ratings provided for the individual items of the dimension. Findings suggest that users rate system responsiveness and reliability above other service quality dimensions. Moderating influences in SERVPERF include gender and pressure to perform. We conclude that (1) companies that provide client services and (2) designers of information systems that interface with users should emphasize responsiveness and reliability; in cases of limited user resources, responsiveness and reliability should be emphasized over other SERVPERF performance dimensions. We also conclude that the use of SERVPERF is nuanced and that moderating influences may affect measurement results.

**Keywords:** Service Quality, SERVQUAL, SERVPERF

## Introduction

---

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact [Publisher@InformingScience.org](mailto:Publisher@InformingScience.org) to request redistribution permission.

As the importance and size of the service sector of the global economy grows, the study of services and innovation are becoming increasingly important. Services are distributed regionally, nationally, and globally and are increasingly becoming a larger portion of many organizations' revenue streams; knowledge intensive business services aimed at enhancing performance require reli-

able methods of measurement, assessment, and improvement (Spohrer & Maglio, 2008). With the aim of sustaining long term relationships with their customers, many businesses have changed their strategic focus to emphasize customer retention (Peng & Wang 2006). Preserving their long term customer relationships requires that these businesses both measure and appropriately adjust the quality of their customer service. Service quality is a major influence on customer satisfaction as customers buy products or services and on whether they continue to do so. As a result, accurate and reliable instruments that assess service quality are of interest to companies whose revenues come in whole or part from service delivery. Currently the most popular and ubiquitous service quality instrument is SERVQUAL.

SERVQUAL is based on the proposition that service quality can be measured as the gap between the service that customers expect and the performance they perceive to have received. Respondents rate their expectations of service from an excellent organization, and then rate the performance they perceive they received from a specific organization. Service quality is calculated as the difference in the two scores where better service quality results in a smaller gap (Landrum, Prybutok, Kappelman, & Zhang, 2008). Although service quality can be evaluated and measured using SERVQUAL, which measures seven service quality dimensions, it also can be measured by its SERVPERF subset, which employs a performance only approach with five dimensions of customers' perceptions of service provider performance.

While both SERVQUAL and SERVPERF provide measures for the same five dimensions of the service quality construct, there exists little published information about the relative importance of each dimension. To supplement information about and improve understanding of the service quality construct, this paper presents results of a study that examined the service quality performance of an electronic library information system (IS). Using the library information system in an engineering research environment, this paper investigates the service quality perceptions of professional system users and reports 1) the relative importance of each of the five SERVPERF dimensions, and 2) the moderating effect of each dimension's importance in the opinion of the system user.

The remainder of this paper is organized as follows: first, we discuss the history of the SERVQUAL instrument and the SERVPERF performance-only subset of SERVQUAL. Next, we discuss the methodology, the two respondent groups, the measures we used, and the data collected. We then discuss the relative importance of each SERVPERF dimension in a series of tables for each of the five dimensions, along with a summary of the findings that show the results we detected in the data. Finally, in the conclusion we discuss our findings and their implications for the SERVQUAL instrument and future service quality research.

## Literature Review

As the service sector of the global economy grows, the study of services and innovation are becoming increasingly important. Service products distributed regionally, nationally, and globally have become larger portions of company revenue streams; knowledge-intensive business services aimed at enhancing performance require reliable methods of measurement, assessment, and improvement (Spohrer & Maglio, 2008). As a result, accurate and reliable instruments that assess service quality are of great interest to companies whose revenues come from service delivery. Perhaps the most popular and widely used service quality instrument is SERVQUAL.

### **Service Quality**

In 1988 Parasuraman, Zeithaml, and Berry developed a generic instrument called SERVQUAL to measure service quality based on input from focus groups. Although SERVQUAL was developed within the marketing sector, it also is used in a variety of organizational settings, including

libraries and information centers (Kettinger & Lee, 1994; Nitecki, 1996). Since 1988 Parasuraman, Zeithaml, and Berry have made numerous changes to SERVQUAL, some in response to problems identified by other researchers. For instance, in 1994 they reported on three different SERVQUAL formats; they recommended that researchers use a format that separated customer expectation scores into tolerance zones.

Researchers have continued to use SERVQUAL instruments. In 1997, Van Dyke, Kappelman, and Prybutok employed SERVQUAL in an IS context, while in 2002 Banwet and Datta measured IT service quality in a library service, as did Landrum and Prybutok in 2004. Still, some researchers question the appropriateness of using SERVQUAL in an IS or IT context; others disagree about whether the service quality should be the difference between expected and perceived service. Parasuraman, Zeithaml, and Berry (1988) stated that since service quality depends on the relationship of customer expectations with customer perceptions, it is appropriate to calculate service quality by subtracting expected from perceived service. One then achieves an overall measure of service quality by averaging the scores of all items (Brown, Churchill, & Peter, 1992). However, this procedure gives also rise to two issues: the first is disagreement over what really is being measured in SERVQUAL with expectations and the second is the problematic nature of the resulting difference scores.

These two issues are resolved if one follows Cronin and Taylor (1992), and Teas (1993), who recommended that expectation ratings be eliminated altogether. In addition, Liljander (1994) states that there is more support for performance only models than for the disconfirmation model of service quality. Bolton and Drew (1991) stated that assessments of overall service quality are affected only by perceptions of performance levels. They suggested that direct measures of disconfirmation are more important than expectations. Boulding, Kalra, Staelin, and Zeithaml (1993) also suggested that perceptions alone influence overall service quality.

Furthermore, other studies suggested that SERVQUAL has unstable dimensions. For example, Jiang, Klein, and Carr (2002) used four dimensions in their study, while Landrum and Prybutok (2004) used five. Nitecki (1996) proposed a three-dimensional SERVQUAL model, as opposed the five dimensions proposed by Zeithaml, Parasuraman, and Berry in 1990. As we have noted, these issues are all resolved if customer expectations are eliminated from the model.

The performance only approach to service quality utilizes the five of the seven SERVQUAL dimensions—the five performance dimensions. Cronin and Taylor (1992), called this performance only subset instrument SERVPERF. When Cronin and Taylor (1992) compared SERVPERF to SERVQUAL, their results supported the dissenters: performance scores alone account for more variation in service quality than performance minus expectations. Performance alone provides better predictive validity than SERVQUAL which is gap-based (Brady, Cronin, & Brand, 2002; Cronin & Taylor 1992) and other studies show that performance scores alone exhibit better reliability and validity than difference scores (Babakus & Boller 1992; Brady et al., 2002; Landrum & Prybutok 2004; Landrum, Prybutok, Strutton, & Zhang, 2008). Based upon these findings, we used only performance scores to perform analysis on the five SERVQUAL service quality dimensions. We next discuss adapting the SERVQUAL instrument to library information services because the test facilities were designated by the US Army Corps of Engineers as “libraries.”

Cook and Thompson (2000) investigated the reliability and validity of SERVQUAL instrument in the context of library service. They found that SERVQUAL displayed three responsive dimensions, rather than the five dimensions originally proposed by Parasuraman et al. (1988). As a result, they concluded that responsive, empathy, and assurance dimensions overlapped in this particular service domain. Nitecki and Hernon (2000) used SERVQUAL to assess library services at Yale University and found that among the five dimensions of SERVQUAL, respondents consid-

ered reliability the most important and empathy least important among the five quality dimensions (Landrum, Prybutok, Kappelman, & Zhang, 2008).

## Measures

### ***SERVQUAL Instrument and its Constructs***

The SERVQUAL instrument used in this study is based on a model of library success that includes the following seven dimensions: service quality, system quality, information quality, user involvement, usefulness, user self-sufficiency, and user satisfaction (Parasuraman et al., 1994). SERVPERF, a SERVQUAL performance only instrument subset of SERVQUAL, uses only the first five of the seven SERVQUAL dimensions – measuring only customers' perceptions of service provider performance. Although additional dimensions, such as individual impact, work group impact, and organizational impact, have been identified as influencing system success (Myers, Kappelman, & Prybutok, 1997), these lie outside the scope of the present study. The SERVQUAL and SERVPERF versions used in this study were tested and validated over a ten year period with a series of studies since its refinement in 1994 (Landrum & Prybutok, 2004; Landrum, Prybutok, Kappelman, & Zhang, 2008; Landrum, Prybutok, Strutton, & Zhang, 2008).

Although the SERVQUAL instrument is ubiquitously employed, it has received heavy criticism from both a theoretical and practical perspective. The issues questioned include the use of gap scores, the overlap among the five dimensions, poor predictive and convergent validity, the ambiguous definition of the “expectation” construct, and unstable dimensionality (Babakus & Boller 1992; Carman, 1990; Brown, Churchill, & Peter, 1992; Van Dyke, Prybutok, & Kappelman, 1999). By discarding the expectations portion in the SERVQUAL model, Cronin and Taylor (1992) justify the SERVPERF or performance only instrument in place of the gap measurement approach. In addition, they show that the SERVPERF instrument empirically outperforms the SERVQUAL scale across several industries. As a result of these issues, the performance only measures are used and suggested by many scholars in various industries (Gilbert, Veloutsou, Goode, & Moutinho, 2004; Keillor, Hult, & Kandemir, 2004; Law, Hui, & Zhao, 2004; Parasuraman et al., 1994; Van Dyke et al., 1997).

We selected a library information system and adapted Parasuraman et al.'s 1994 version of SERVQUAL with 21 questions in 5 dimensions of service performance to study it (Appendix, Section 1). We made minimal wording changes to the original content of the instrument; although we did measure the customer expectations, we did not use them in this study because we focused on the performance (SERVPERF). This focus was deemed appropriate because the intent was to examine the relative importance of the performance dimensions.

The 21 questions are distributed among the 5 SERVPERF performance dimensions as follows: 1) five items are used to measure tangibles (questions 1-5) – tangibles refer to physical facilities, equipment, and personnel; 2) five items are used to measure reliability (questions 6-10) – reliability refers to the ability of a firm to perform promised service dependably and accurately; 3) four items are used to measure responsiveness (questions 11-14) – responsiveness is the willingness to help customers and provide prompt service; 4) four items are used to measure assurance (questions 15-18) – assurance is the ability of employees to inspire trust and confidence in customers; and 5) three items are used to measure empathy (questions 19-21) – empathy refers to the amount of caring and individualized attention provided to customers. We also collected point data from each respondent indicating the perceived relative importance of each of the five SERVPERF dimensions; respondents were instructed to allocate 100 points among the five dimensions (Appendix, Section 2).

## ***Research Methodology and Data Collection***

This research seeks to answer the following question: Do users (study respondents) assign importance ratings for each of the five SERVPERF dimensions that differ from their allocation of importance points to each of the same five dimensions? This question addresses what we believe may be a gap in the research knowledge about the SERVQUAL and SERVPERF instruments and provides possible prescriptive nuances as to their use. There exist two compelling reasons for this belief: First, the literature contains multiple instances where researchers have chosen three, four, and five of the SERVPERF dimensions to investigate performance service quality. These choices indicate possible disagreement among researchers as to the relevance and importance of the SERVPERF dimensions. Second, Nitecki and Herson (2000) reported that library system users ranked the SERVPERF reliability dimension most important and the empathy dimension least important. These rankings indicate disagreement among users about the relative importance of the SERVPERF dimensions. As a result, we wonder whether some SERVPERF dimensions are more or less relevant to service quality than others, and if so, which ones are the most important to library information system users. This paper contributes to understanding the relative dimension importance issues that we perceive in SERVPERF service quality measures.

To answer the research question, we used a mail survey to collect data from a population of library users at the U.S. Army Corps of Engineers Waterways Experiment Station (WES), in Vicksburg, MS, as well as the U.S. Army Corps of Engineers Little Rock District Office in Little Rock, AR. Both facilities sites have similar missions and types of knowledge-worker customers, both are information service centers, and both are called “research libraries.” The research areas of the knowledge-worker customers encompass environmental studies, coastal engineering, soils analysis, concrete and dam structure analysis, and hydraulics. Because the customers consist primarily of government engineers, scientists, and support staff, the customer base at both facilities is nearly identical.

Based on the information supplied to the author by the Headquarters Office of the U.S. Army Corps of Engineers, the number of federal employees at Little Rock is 765 and the number of employees at Vicksburg is 1,228. Of the total number of employees, about 45% of the work forces are engineers at Vicksburg and about 28% of the work forces are engineers at Little Rock. At Little Rock about 71% of the work forces are men, and at Vicksburg, about 69% of the work forces are men. The mean age of the worker at Little Rock is 44; the mean age of the worker at Vicksburg is 45. The difference between the Vicksburg and Little Rock is primarily the size of the population.

The mission of both libraries is to provide support for both scientific and legal research related to U.S. Army Corps of Engineers projects. The work is typical of most corporate or special libraries. Library use is voluntary at both sites. There are no direct charges for using the library, except when performing lengthy, fee-based database searches. However, most users are aware there are costs associated with providing library service because each laboratory or agency is assessed an overhead fee for library operation. Library services offered at both sites include checking books in and out for customers, ordering books and article reprints for customers, requesting interlibrary loans for customers, routing journals, and providing search assistance in various online databases.

Survey respondents were chosen randomly from user lists at each site. We distributed and received the questionnaires anonymously in sealed envelopes through the chief librarian at each library. Respondents were asked not to write their names on the questionnaire and were assured that all responses would be treated anonymously. The chief librarians gathered all returned survey forms and mailed them to the authors. While these facilities had not previously participated in a service quality survey, the director of the library service at the facilities was interested and willing to participate in the survey because he deemed the feedback valuable.

Respondents evaluated the perceived performance of the IS for each SERVPERF item on a seven-point scale ranging from low to high. We collected a total of 385 usable responses with a response rate of 37%; 70% of the respondents were men and 68% engineers or scientists. The percentages closely reflect the demographics of the study sites. More than 75% of the respondents used the center frequently and at least 77% indicated they relied on center staff frequently when looking for information. The respondents' knowledge of and experience with the research libraries qualified them to assess IS service performance. The survey instrument is shown in the Appendix.

## Data Analysis

Using the SPSS statistics software package to analyze data collected from the SERVPERF instrument shown in the Appendix, we examined user attitudes regarding the SERVPERF service quality dimensions. We asked the respondents to gauge the importance of each of the five dimensions two ways: first by allocating importance points to each dimension, and second by assigning an importance ranking to each dimension. By comparing results of the two measures, we can detect differences between how users allocate points with constraints (the respondents cannot allocate more or less than 100 total points to the five dimensions) and how users rank each dimension without constraints (the respondents can assign any rank on the 7-item Likert scale to any dimension).

Table 1 shows that, on average, respondents assigned most points to reliability, followed by responsiveness, assurance, empathy, lastly tangibility. Table 2 shows the descriptive statistics of the five SERVPERF dimensions based on Likert-scale ratings. The Table 2 data will be split into 3 equal groups, with only the high and low polar thirds analyzed for consistency, as indicated in Tables 3 through 12.

<i>Dimension points</i>	<i>Mean</i>	<i>Standard Deviation</i>
Tangibility points	12.29	7.56
<b>Reliability points</b>	<b>32.43</b>	<b>15.14</b>
<b>Responsiveness points</b>	<b>22.09</b>	<b>8.29</b>
Assurance points	17.88	7.59
Empathy points	15.27	8.97

<i>Dimension rankings</i>	<i>Mean</i>	<i>Standard Deviation</i>
Tangibility	5.73	0.92
Reliability	5.91	0.96
Responsiveness	6.16	0.96
Assurance	5.96	0.94
Empathy	5.95	1.00

### ***Analysis of the Five SERVPERF Dimensions with Polar Extremes***

We subsequently performed data analysis with the polar groups for each of the five SERVPERF dimensions—the five groups of respondents that ranked the five SERVPERF dimensions high and the five groups of respondents that ranked the five SERVPERF dimensions low—yielding ten total groups or five polar pairs. Using the five polar pairs, we performed a series of five tests as follows: we compared the five high and low polar pairs with the importance points these same respondents allocated to 1) the tangibility dimension (Tables 3 and 4), 2) the reliability dimension (Tables 5 and 6), 3) the responsiveness dimension (Tables 7 and 8), 4) the assurance dimension (Tables 9 and 10), and 5) the empathy dimension (Tables 11 and 12).

We used this method of polar extremes to emphasize the differences among the respondents, as indicated by their own opinions of each dimension's relative importance—based on the number of points they allocated to each dimension.

Our purpose is to examine differences between two user groups, and using the polar extremes maximizes the differences among the users and, thereby, increases the opportunity to gain insight into the user's service preferences. As described by Hair, Black, Babin, Anderson, and Tatham, (2006, p. 288), “The polar extreme approach involves comparing only the extreme two groups and excluding the middle group from the discriminant analysis.” This approach can reveal differences that are not as prominent with other techniques, such as regression analysis or discriminant analysis with the data split equally. We believe that the differences between the two extreme groups are more salient than comparing equally split groups; as a result of the polar extreme data split in conjunction with the use of comparing the means of these two extreme groups we retained all extreme responses including the outliers in our data. We also believe that this technique provides richer information to a company for the design of its customer service function (please see SERVPERF in the Appendix, Section 2, Questions 1 through 5). For instance, reliability points (Section 2, Question 2) allocated by respondents in the two extreme groups can be used to analyze and target service differentially to the particular expectations of customers in each these two separate groups. Therefore, we compared the means differences along the five SERVPERF dimensions for the two polar extreme (high and low) groups and then conducted t-tests to compare the mean-response differences along the five SERVPERF dimensions.

### ***Analysis of SERVPERF Dimension 1: Tangibility***

Beginning with the first dimension, tangibility, we split the data into three equal data sets based on the tangibility-points frequencies (see Table 3). The high group contained the upper 33.33 percentile of the data and the low group contained the lower 33.3 percentile of the data points; we perform this same procedure for the remaining SERVPERF dimensions, as well.

Table 3 shows the first set of descriptive statistics for the two polar extreme groups of Dimension 1. For instance, in dimension row 1, group 1 respondents allocated an average of 82 points to the tangibility dimension, while group 3 respondents allocated an average of 80 points to the tangibility dimension. Table 4 shows that the t-test results from comparing the tangibility points means show statistically significant differences between the high and low groups on tangibility, as expected, and responsiveness scales, which was not expected. The tangibility differences observed are consistent with our SERVPERF expectations; the responsiveness differences are inconsistent with our SERVPERF expectations.

**Table 3. Dimension 1 (Tangibility). Comparison of the 5 polar pairs with the importance points allocated to the tangibility dimension. We perform t-tests to compare the means for significant differences in Table 4.**

<i>Dimension</i>	<i>Tangible Points Allocated Group</i>	<i>N</i>	<i>Mean</i>	<i>Standard Deviation</i>
Tangibility	1	82	5.54	0.94
	3	80	6.01	0.91
Reliability	1	76	6.02	0.93
	3	73	5.79	1.08
Responsiveness	1	85	6.34	0.86
	3	81	5.96	1.26
Assurance	1	79	5.95	0.92
	3	76	5.87	1.18
Empathy	1	82	6.04	0.97
	3	76	5.75	1.24

**Table 4. Dimension 1 (Tangibility). T-test results from comparing the tangibility points means show statistically significant differences between the high and low groups on tangibility and responsiveness scales. The characters “\*\*” indicate statistical significance for the listed dimension.**

<i>Dimension</i>	<i>t-test</i>	<i>Significance</i>	<i>Mean Difference</i>
<b>Tangibility **</b>	<b>-3.68</b>	<b>0.00</b>	<b>-0.47</b>
Reliability	1.418	0.16	0.23
<b>Responsiveness **</b>	<b>2.19</b>	<b>0.03</b>	<b>0.38</b>
Assurance	0.45	0.65	0.08
Empathy	1.71	0.09	0.30

### **Analysis of SERVPERF Dimension 2: Reliability**

We used the same method to create two polar extreme groups based on the points allocated to the reliability dimension. Table 5 shows the descriptive statistics for the dimensions based on these two groups. For instance, in Table 5 dimension row 2, group 1 respondents allocated an average of 95 points to the reliability dimension, while group 3 respondents allocated an average of 70 points to the reliability dimension. The Table 6 t-test results from comparing the reliability points means show unexpected but significant differences between the high and low groups on tangibility and assurance scales, but no differences on reliability, as expected. The tangibility and assurance differences observed are inconsistent with our SERVPERF expectations; the absence of differences on the reliability scale are also inconsistent with our a priori SERVPERF expectations.

**Table 5. Dimension 2 (Reliability). Comparison of the 5 polar pairs with the importance points allocated to the reliability dimension. We perform t-tests to compare the means for significant differences in Table 6.**

<i>Dimension</i>	<i>Polar Reliability Points Allocated Group</i>	<i>N</i>	<i>Mean</i>	<i>Standard Deviation</i>
Tangibility	1	103	5.93	0.86
	3	69	5.57	0.98
Reliability	1	95	5.95	1.05
	3	70	5.81	0.94
Responsiveness	1	108	6.17	1.02
	3	76	6.12	1.02
Assurance	1	99	6.13	0.92
	3	74	5.83	0.97
Empathy	1	105	6.06	1.08
	3	72	5.80	1.04

**Table 6. Dimension 2 (Reliability). T-test results from comparing the reliability points means show statistically significant differences between the high and low groups on tangibility and assurance scales but no differences on reliability. The characters “\*\*” indicate statistical significance.**

<i>Dimension</i>	<i>t-test</i>	<i>2-Tailed Significance</i>	<i>Mean Difference</i>
<b>Tangibility **</b>	<b>2.59</b>	<b>0.01</b>	<b>0.37</b>
Reliability	0.89	0.37	0.14
Responsiveness	0.32	0.75	0.05
<b>Assurance **</b>	<b>2.02</b>	<b>0.05</b>	<b>0.29</b>
Empathy	1.65	0.10	0.27

### **Analysis of SERVPERF Dimension 3: Responsiveness**

We used the same method to create two polar extreme groups based on the points allocated to the responsiveness dimension. Table 7 shows the descriptive statistics for the dimensions based on these two groups. Table 8 shows that the two polar extreme groups created based on the point allocation to responsiveness dimension differ in the item rating means for the reliability and empathy dimensions. For instance, in Table 7 dimension row 3, group 1 respondents allocated an average of 40 points to the responsiveness dimension, while group 3 respondents allocated an average of 142 points to the responsiveness dimension. Table 8 t-test results from comparing the responsiveness points means show unexpected significant differences between the high and low groups on reliability and empathy scales, but no differences on responsiveness, as expected. The reliability and empathy differences observed are inconsistent with our SERVPERF expectations; the absence of differences on the responsiveness scale also is inconsistent with our SERVPERF expectations.

**Table 7 Dimension 3 (Responsiveness). Comparison of the 5 polar pairs with the importance points allocated to the responsiveness dimension. We perform t-tests to compare the means for significant differences in Table 8.**

<i>Dimension</i>	<i>Responsiveness Points Allocated Group</i>	<i>N</i>	<i>Mean</i>	<i>Standard Deviation</i>
Tangibility	1	39.00	5.70	1.04
	3	135.00	5.62	0.90
Reliability	1	35.00	5.46	1.24
	3	126.00	5.92	0.96
Responsiveness	1	40.00	5.93	1.31
	3	142.00	6.22	0.92
Assurance	1	39.00	5.64	1.11
	3	129.00	5.96	0.89
Empathy	1	37.00	5.57	1.30
	3	136.00	5.95	0.97

**Table 8. Dimension 3 (Responsiveness) T-test results from comparing the responsiveness points means show statistically significant differences between the high and low groups on reliability and empathy scales but no differences on responsiveness. The characters “\*\*” indicate statistical significance.**

<i>Dimension</i>	<i>t-test</i>	<i>2-Tailed Significance</i>	<i>Mean Difference</i>
Tangibility	0.48	0.63	0.08
<b>Reliability **</b>	<b>-2.34</b>	<b>0.02</b>	<b>-0.46</b>
Responsiveness	-1.59	0.11	-0.29
Assurance	-1.85	0.07	-0.32
<b>Empathy **</b>	<b>-1.94</b>	<b>0.05</b>	<b>-0.38</b>

### **Analysis of SERVPERF Dimension 4: Assurance**

We used the same method to create two polar extreme groups based on the points allocated to the assurance dimension. Table 9 shows the descriptive statistics for the dimensions based on these two groups. Table 10 shows that the two polar extreme groups created based on the point allocation to assurance dimension differ in the mean item ratings for the reliability, assurance, and empathy dimensions. For instance, in Table 9 dimension row 4, group 1 respondents allocated an average of 89 points to the assurance dimension, while group 3 respondents allocated an average of 64 points to the assurance dimension. In Table 10 the t-test results from comparing the assurance points means show unexpected significant differences between the high and low groups on reliability and empathy scales, as well as expected differences on assurance. The reliability and

empathy differences observed are inconsistent with our SERVPERF expectations; the differences on the assurance scale are consistent with our SERVPERF expectations.

**Table 9 Dimension 4 (Assurance). Comparison of the 5 polar pairs with the importance points allocated to the assurance dimension. We perform t-tests to compare the means for significant differences in Table 10.**

<i>Dimension</i>	<i>Polar Assurance Points Allocated Group</i>	<i>N</i>	<i>Mean</i>	<i>Standard Deviation</i>
Tangibility	1	87	5.70	0.99
	3	68	5.71	0.87
Reliability	1	84	5.67	1.12
	3	59	6.10	0.93
Responsiveness	1	96	5.99	1.22
	3	71	6.24	0.82
Assurance	1	89	5.68	1.05
	3	64	6.16	0.82
Empathy	1	91	5.66	1.18
	3	70	6.15	0.87

**Table 10. Dimension 4 (Assurance) T-test results from comparing the assurance points means show statistically significant differences between the high and low groups on reliability and empathy scales, but no differences on assurance. The characters “\*\*” indicate statistical significance.**

<i>Dimension</i>	<i>t-test</i>	<i>2-Tailed Significance</i>	<i>Mean Difference</i>
Tangibility	-0.10	0.92	-0.02
<b>Reliability **</b>	<b>-2.41</b>	<b>0.02</b>	<b>-0.43</b>
Responsiveness	-1.50	0.14	-0.25
<b>Assurance **</b>	<b>-3.04</b>	<b>0.00</b>	<b>-0.48</b>
<b>Empathy **</b>	<b>-2.93</b>	<b>0.00</b>	<b>-0.49</b>

### **Analysis of SERVPERF Dimension 5: Empathy**

We used the same method to create two polar extreme groups based on the points allocated to the empathy dimension. Table 11 shows the descriptive statistics for the dimensions based on these two groups. Table 12 shows that the two polar extreme groups created based on the point allocation to empathy dimension differ in the mean item ratings for only the reliability dimension. For example, in Table 11 dimension row 5, group 1 respondents allocated an average of 47 points to the empathy dimension, while group 3 respondents allocated an average of 31 points to the empathy dimension. In Table 12 the t-test results from comparing the empathy points means show unexpected significant differences between the high and low groups on assurance and expected differences on the empathy scale.. The assurance differences observed are inconsistent with our

SERVPERF expectations; the differences on the empathy scale are consistent with our SERVPERF expectations.

**Table 11 Dimension 5 (Empathy). Comparison of the 5 polar pairs with the importance points allocated to the empathy dimension. We perform t-tests to compare the means for significant differences in Table 12.**

<i>Dimension</i>	<i>Polar Empathy Points Allocated Group</i>	<i>N</i>	<i>Mean</i>	<i>Standard Deviation</i>
Tangibility	1	46	5.62	0.96
	3	29	5.77	0.86
Reliability	1	45	5.84	1.01
	3	27	5.89	1.10
Responsiveness	1	50	6.09	1.10
	3	33	6.38	0.68
Assurance	1	46	5.76	1.06
	3	30	6.20	0.78
Empathy	1	47	5.78	1.07
	3	31	6.31	0.90

**Table 12. Dimension 5 (Empathy). T-test results from comparing the empathy points means show statistically significant differences between the high and low groups on assurance and empathy scales. The characters “\*\*\*” indicate statistical significance.**

<i>Dimension</i>	<i>t-test</i>	<i>2-Tailed Significance</i>	<i>Mean Difference</i>
Tangibility	-0.69	0.50	-0.15
Reliability	-0.21	0.83	-0.05
Responsiveness	-1.35	0.18	-0.29
Assurance **	<b>-1.96</b>	<b>0.05</b>	<b>-0.44</b>
<b>Empathy **</b>	<b>-2.26</b>	<b>0.03</b>	<b>-0.52</b>

## Discussion

Table 13 summarizes the previously presented t-tests, indicating where significant differences exist (cells containing “\*\*\*”) between the two groups of respondents, indicating where our findings are consistent with SERVPERF (cells containing “C”), and indicating where our findings were inconsistent with SERVPERF (cells containing “I”). If we were to assume that our data collection is flawless and SERVPERF is accurately measuring what it is supposed to be measuring, then we would expect to see only five “C \*\*\*” cells on the descending diagonal of Table 13.

**Table 13. This table summarizes the previously presented t-tests, indicating where significant differences exist (cells containing “\*\*”) between the two groups of respondents, indicating where our findings are consistent with SERVPERF (cells containing “C”), and indicating where our findings were inconsistent with SERVPERF (cells containing “I”).**

	Tangibil- ity	Reliability	Responsiveness	Assurance	Empathy
Tangibility points	<b>C **</b>		I **		
Reliability points		<b>I- no signif</b>		I **	
Responsiveness points	I **	I **	<b>I-no signif</b>		
Assurance points		I **		<b>C **</b>	I **
Empathy points				I **	<b>C **</b>

We find that our results are consistent with SERVPERF only in the respondent point allocations for tangibility, assurance, and empathy (marked “C \*\*” along the descending diagonal for “consistency and significance found”). Here, the respondents in the high polar and low polar groups behave differently, as expected; i.e., respondents that rate a dimension low (allocate few points) may be less demanding of service quality in that area, as opposed to respondents who rate the dimension high (allocate more points), who may be more demanding in their SERVPERF responses. Thus, results the tangibility, assurance, and empathy dimensions along the diagonal are consistent with expected results.

The polar means for the reliability and responsiveness dimensions exhibit no significant differences and are not consistent with *a priori* expectations. Even though respondents in both groups ranked the importance of reliability and responsiveness differently, they still allocated a comparable number of points to both dimensions. In other words, both respondent groups consider reliability and responsiveness equally vital to perform their research, regardless of their intellectual view of the tasks.

Based on the average points allocated to each dimension, we would anticipate that differences would exist in the most important dimensions. However, Table 1 indicates that the three dimensions with the lowest point allocations show consistency and the two dimensions that were related most important do not. One could reason that there exists only a weak association between the respondents’ perceptions (measured by point allocations) and their responses to the individual SERVPERF items. Or, perhaps the respondents do not allocate importance points in the same way they respond to SERVPERF reliability and responsiveness service quality questions.

However, it is also possible that a previously unaccounted-for moderating effect is responsible for the respondents’ inconsistent responses—a shared environmental characteristic experienced by all the professionals at both US Army engineering facilities. We note that three of the service quality dimensions (tangibility, assurance, and empathy) are more subjective characteristics of a professional work environment, while reliability and responsiveness are more objective characteristics. In other words, these two metric groups are emotional opposites. In our data, 68% of the respondents were engineers or scientists. In addition, engineering performance (a common metric used to allocate performance rewards) is primarily measured quantitatively and objectively. Also, in our data, 70% of the respondents were men.

In the work environment, Rowlands (2008) found that women were more likely to be affected by work relationships, whereas men care relatively less about others and struggle to perform and

cope with change. As a result, we believe that the two statistically-inconsistent responses reveal a missing, moderating influence on the SERVPERF service quality expectations. We suggest that the missing influence is the moderating effect of pressure to perform on the user's expectations and perceptions of service quality, affecting SERVPERF measurements. We also believe that gender may interact with the moderating stress variable to affect responses on the SERVPERF (and SERVQUAL) measurements.

We offer as a possible explanation that, because the majority of the respondents were male engineers and because professional males characteristically minimize the importance of subjective effects and maximize the importance of objective effects under stress, under conditions of job stress (e.g., performance deadlines, accuracy concerns) the reliability and responsiveness dimensions of service performance were rated as the most important service quality concerns.

In our examination of SERVPERF using a library information system in an engineering research environment, we reason that, if the SERVPERF consistently and accurately measures each of these dimensions, then the service quality received by system users will align exactly with the number of importance points they allocate to each of these dimensions. In other words, if the number of importance points allocated by the respondents before measuring service quality was consistent with what is really measured by SERVPERF, then the highlighted cells should plot *only* along the descending diagonal of Table 13. But they do not, which suggests that the SERVPERF instrument contains inconsistencies in its measurement methodology. We noted earlier that previous studies disagreed in the number of dimensions varying from three to five in recent SERVQUAL research. Furthermore, we find that using the average rating of items in SERVPERF will not necessarily agree with the importance a user places on a dimension. Our findings raise questions as to what SERVPERF is really measuring.

## Conclusion

The results of this study suggest that the five dimensions of SERVPERF, which is a subset of SERVQUAL, are both dichotomous and hierarchical, and subject to conditioning based on user perceptions of stress and urgency. In other words, SERVQUAL the instrument is not homogeneous, and some dimensions potentially matter more to users than others. It is possible that such differences are a function of the user's state of mind.

The results suggest that the performance dimension of reliability, which refers to the ability of a firm to perform promised service dependably and accurately, and the dimension of responsiveness, which is the ability to provide prompt service, consistently rank highly in both respondent groups, while the more emotive dimensions of tangibility, assurance, and empathy vary in importance depending on the user's state of mind and the circumstances. When experiencing job-related pressure, users may well demand speed and reliability because this is what their deadline requires. Stress melts some of the veneer of collegiality and with it the need for tangibility, assurance, and empathy, which may become relegated to a relatively less important role in the work place. Our results are consistent with Nitecki and Hernon (2000), who also used SERVQUAL to assess library services at Yale University, where respondents ranked reliability the most important of the five SERVPERF dimensions, and ranked empathy, one of the relationship dimensions, to be the least important.

The main difference between our study and Nitecki and Hernon's is that our data analysis provides additional insights into service measurement issues, where we examine differences between the polar means of two groups of users. In some circumstances, polar means analysis can reveal differences in the data that are not as evident as with other techniques, such as regression analysis or discriminant analysis, where the data are split equally.

This research has raised some important new questions not addressed by earlier studies on service quality. Namely, is there a missing moderating variable or variables (e.g., user mental state, stress level, or perceived job security), and how do they interact with gender and / or occupation in the dimensions of service quality? We believe this question is applicable to all service quality areas, including IT service areas, because most jobs are subject to conditions of stress and most jobs are populated by a mix of males and females. What we may in fact be studying is two levels or planes of service quality: nice-to-have services and essential services. These findings suggest that, in high-stress environments, the more objective performance characteristics of reliability and responsiveness rule.

If our findings are correct, individuals who research service levels with SERVQUAL and SERVPERF have new limitations to consider. 1) If they accept our findings and also accept that SERVQUAL accurately measures service quality, the findings suggest that companies should focus relatively more of their service quality efforts on the reliability and responsiveness dimensions. Given limited resources and economic pressures, a company that recognizes the relative importance of each dimension can more efficiently allocate resources and focus service priorities. 2) If they accept our findings, but adjust their data analysis of SERVQUAL results to account for ranked dimensions whose outcomes are affected by the user's state of mind, then their choices are not as clear. They are faced with calibrating SERVQUAL to account for the moderating effect of user mental state or they can choose another service quality instrument. 3) Finally, they can simply ignore these findings and proceed as before.

As service industries and service jobs continue to expand in importance throughout the global economy, measuring service quality quickly, reliably, and accurately will mature as an increasingly important task for service providers. If SERVQUAL and its performance subset SERVPERF are to be relied upon by services companies for their survival, then understanding the nuances of what these instruments are measuring has significant financial implications.

## References

- Babakus, E., & Boller, G. W. (1992). An empirical assessment of the SERVQUAL Scale. *Journal of Business Research, 24*(3), 253-268.
- Banwet, D. K., & Datta, K. (2002). Effect of service quality on post visit intentions over time. The case of a library. *Total Quality Management, 13*(4), 537-546.
- Bolton, R. N., & Drew, J. H. (1991). A multistage model of customers' assessments of service quality and value. *Journal of Consumer Research, 17*(4), 375-384.
- Boulding, W., Kalra, A., Staelin, R., & Zeithaml, V. (1993). A dynamic process model of service quality: From expectations to behavioral intentions. *Journal of Marketing Research, 30*, 7-27.
- Brady, M. K., Jr., Cronin, J., Jr., & Brand, R. R. (2002). Performance-only measurement of service quality: A replication and extension. *Journal of Business Research, 55*, 27-31.
- Brown, T. J., Churchill, G. A., Jr., & Peter, J. P. (1992). *Improving the measurement of service quality* (Working Paper No. 92-4). Madison, WI: A.C. Nielsen Center for Marketing Research.
- Carman, J. M. (1990). Consumer perceptions of service quality: An assessment of SERVQUAL dimensions. *Journal of Retailing, 66*(1), 33-55.
- Cook, C., & Thompson, B. (2000). Reliability and validity of SERVQUAL scores used to evaluate perceptions of library service quality. *Journal of Academic Librarianship, 26*(4), 248-258.
- Cronin, J. J., Jr., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing, 56*, 55-68.
- Gilbert, G. R., Veloutsou, C., Goode, M. M. H., & Moutinho, L. (2004). Measuring customer satisfaction in the fast food industry: A cross-national approach. *The Journal of Services Marketing, 18*(5), 371-383.

## Measuring IS System Service Quality with SERVQUAL

- Hair, F. H., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis*. Upper Saddle River, NJ: Pearson Prentice-Hall.
- Jiang, J. J., Klein, G., & Carr, C. L. (2002). Measuring information system service quality: SERVQUAL from the other side. *MIS Quarterly*, 26(2), 145-166.
- Keillor, B. D., Hult, G. T. M., & Kandemir, D. (2004). A study of the service encounter in eight countries. *Journal of International Marketing*, 12(1), 9-35.
- Kettinger W. J., & Lee, C. C. (1994). Perceived service quality and user satisfaction with the information services function. *Decision Sciences*, 25(6), 737-766.
- Landrum, H., & Prybutok, V. R. (2004). A service quality and success model for the information service industry. *European Journal of Operational Research*, 156(2), 628.
- Landrum, H., Prybutok, V. R. Kappelman, L. A., & Zhang, X. (2008). SERVCESS: A parsimonious instrument to measure service quality and information system success. *The Quality Management Journal*, 15(3), 17-25
- Landrum, H., Prybutok, V. R., Strutton, D., & Zhang, X. (2008). Examining the merits of usefulness versus use in an information service quality and information system success web-based model. *Information Resources Management Journal*, 21(2), 1-17.
- Law, A. K. Y., Hui, Y. V., & Zhao, X. (2004). Modeling repurchase frequency and customer satisfaction for fast food outlets. *The International Journal of Quality & Reliability Management*, 21(5), 545-563.
- Liljander, V. (1994). Modeling perceived service quality using different comparison standards. *Journal of Consumer Satisfaction*, 7, 126-142.
- Myers, B. L., Kappelman, L. A., & Prybutok, V. R. (1997). A comprehensive model for assessing the quality and productivity of the information systems function: Toward a theory for information systems assessment. *Information Resources Management Journal*, 10(1), 6-25.
- Nitecki, D. (1996). Changing the concept and measure of service quality in academic libraries, *Journal of Academic Librarianship*, 2(3), 181-190.
- Nitecki, D. A., & Herson, P. (2000). Measuring service quality at Yale University's libraries. *Journal of Academic Librarianship*, 26(4), 259-273.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions. *Journal of Retailing*, 64(1), 12.
- Parasuraman, A, Zeithaml, V. A., & Berry, L. L. (1994). Alternative scales for measuring service quality: A comparative assessment based on psychometric and diagnostic criteria. *Journal of Retailing*, 70(3), 201.
- Peng, L. Y., & Wang, Q. (2006). Impact of relationship marketing tactics (RMTs) on switchers and stayers in a competitive service industry. *Journal of Marketing Management*, 22(1,2), 25-59.
- Rowlands, R. (2008). Stress agender. *Occupational Health*, 60(6), 26-27.
- Spohrer, J., & Maglio, P. (2008). The emergence of service science: Toward systematic service innovations to accelerate co-creation of value. *Production and Operations Management*, 17(3), 238-246.
- Teas, R. K. (1993). Expectations, performance evaluation and consumer's perception of quality. *Journal of Marketing*, 57(4), 18-34.
- Van Dyke T. P., Kappelman, L. A., & Prybutok, V. (1997). Measuring information systems service quality: Concerns on the use of the SERVQUAL questionnaire. *MIS Quarterly*, 21(2), 195-208.
- Van Dyke, T. P., Prybutok, V. R., & Kappelman, L. (1999). Cautions on the use of the SERVQUAL measure to assess the quality of information systems services. *Decision Sciences*, 30(3), 877-892.
- Zeithaml, V.A., Parasuraman, A, & Berry, L. L. (1990). *Delivering quality service: Balancing customer perceptions and expectations*. New York, NY: Macmillan.

## Appendix: The SERVPERF Questionnaire

**Section 1. DIRECTIONS:** *We want your impressions about how the Waterways Experiment Station Research Library performs relative to your expectations. For each of the items below, please indicate your perception of the library performance by circling a number. There are no right or wrong answers. If you are unsure how to rate the library's performance, circle the number under "unknown". Please do not omit any feature.*

	How the library performs here.		
	low	high	unknown
When it comes to...			
1. Modern equipment.	1 2 3 4 5 6 7		0
2. Visually appealing facilities.	1 2 3 4 5 6 7		0
3. Neat, professionally appearing staff.	1 2 3 4 5 6 7		0
4. Visually appealing documentation, such as library signs, handouts, and brochures.	1 2 3 4 5 6 7		0
5. Convenient hours of operation.	1 2 3 4 5 6 7		0
6. Providing service as promised.	1 2 3 4 5 6 7		0
7. Dependability in handling users' service problems.	1 2 3 4 5 6 7		0
8. Performing service right the first time.	1 2 3 4 5 6 7		0
9. Providing service at the promised time.	1 2 3 4 5 6 7		0
10. Keeping users informed about when services will be performed.	1 2 3 4 5 6 7		0
11. Prompt service to users.	1 2 3 4 5 6 7		0
12. Willingness to help users.	1 2 3 4 5 6 7		0
13. Readiness to respond to users' requests.	1 2 3 4 5 6 7		0
14. Courteous staff.	1 2 3 4 5 6 7		0
15. Staff who instill confidence in users.	1 2 3 4 5 6 7		0
16. Making users feel secure in their transactions with the library.	1 2 3 4 5 6 7		0
17. Staff who have the knowledge to answer users' questions.	1 2 3 4 5 6 7		0
18. Giving users individual attention.	1 2 3 4 5 6 7		0
19. Staff who have the users' best interests at heart.	1 2 3 4 5 6 7		0
20. Staff who deal with users in a caring fashion.	1 2 3 4 5 6 7		0
21. Staff who understand the needs of users.	1 2 3 4 5 6 7		0

**Section 2. DIRECTIONS:** *Listed below are five features pertaining to libraries and the service they offer. We would like to know how important each of these features is to you when you evaluate a library's quality of service. Please allocate a total of 100 points among the five features according to how important each feature is to you – the more important a feature is to you, the more points you should allocate to it. Please ensure that the points you allocate to the five features add up to 100.*

1. The appearance of the library's physical facilities, equipment, and personnel. \_\_\_\_\_ points
2. The library's ability to perform the promised service dependably and accurately. \_\_\_\_\_ points
3. The library's willingness to help users and provide prompt service. \_\_\_\_\_ points
4. The knowledge and courtesy of the library's staff and their ability to convey trust and confidence. \_\_\_\_\_ points
5. The caring, individualized attention the library provides users. \_\_\_\_\_ points

**Total points allocated: 100 points**

## Biographies



**Hollis T. Landrum** retired as an information systems management specialist with the U.S. Army Corps of Engineers at the Engineer Research and Development Center. He received a doctorate in information science from the University of North Texas in 1999. He has published numerous articles in information and library science. In addition, he has published articles in national music journals and has had several plays produced. Landrum has won numerous awards for his work with the Army Corps of Engineers, including the Commander's Award for Civilian Service, one of the highest civilian awards in the Army.



**Victor R. Prybutok** is a Regents professor of decision sciences in the Information Technology and Decision Sciences Department and is director of the Center for Quality and Productivity in the College of Business Administration at the University of North Texas. He received a doctorate from Drexel University in 1984. He is a Senior member of the American Society for Quality (ASQ) and is an ASQ certified quality engineer, certified quality auditor, and certified quality manager. Journals where his published articles have appeared include *The American Statistician*, *Communications of the ACM*, *Communications in Statistics*, *Data Base*, *Decision Sciences*, *European Journal of Operational Research*, *IEEE Transactions on Engineering Management*, *MIS Quarterly*, *OMEGA: The International Journal of Management Science*, *Operations Research*, and *Quality Management Journal*. In addition, he serves on the editorial board of the *Quality Management Journal*. He can be reached by email at [Prybutok@unt.edu](mailto:Prybutok@unt.edu).



**Xiooni Zhang** is an associate professor of business informatics at the Northern Kentucky University. She received her doctorate in business computer information systems from the University of North Texas in 2001. Her research interests include software volatility, Web usability, e-commerce systems, enterprise systems, and mobile technology. She is a member of the Decision Science Institute and Association for Information System. Her publications appear in *IEEE Transactions on Engineering Management*, *Communication of the ACM*, *Information & Management*, *International Conference of Information Systems*, and other journals. Zhang can be reached by email at [zhangx@nku.edu](mailto:zhangx@nku.edu).



**Daniel A. Peak** is an associate professor in information technology in the Information Technology and Decision Sciences Department, College of Business Administration at the University of North Texas. He received his Ph.D. in 1994 from UNT with majors in Information Systems and in Finance. He also has several degrees in piano performance, and has studied with well-known pianists and artists. He is an editor of the *Journal of IT Cases and Applications Research*. Dr. Peak has more than 20 years of IT consulting and planning experience working for executives of Fortune 500 companies, and has won and directed numerous production projects and research grants. He is a member of the Decision Science Institute and Association for Information Systems and has publications in *Informing Science*, *Information and Management*, *Information Systems Management*, and other journals. Dan can be reached by email at [peak@unt.edu](mailto:peak@unt.edu).