Developing a new key performance index for measuring service quality
Jeh-Nan Pan and Tzu-Chun Kuo
Department of Statistics, National Cheng Kung University, Tainan, Taiwan

Abstract
Purpose – The purpose of this research is to develop a new key performance index (KPI) and its interval estimation for measuring the service quality from customers’ perceptions, since most service quality data follow non-normal distribution.

Design/methodology/approach – Based on the non-normal process capability indices used in manufacturing industries, a new KPI suitable for measuring service quality is developed using Parasuraman’s 5th Gap between customers’ expectation and perception. Moreover, the confidence interval of the proposed KPI is established using the bootstrapping method.

Findings – The quantitative method for measuring the service quality through the new KPI and its interval estimation is illustrated by a realistic example. The results show that the new KPI allows practising managers to evaluate the actual service quality level delivered within each of five SERVQUAL categories and prioritize the possible improvement projects from customers’ perspectives. Moreover, compared with the traditional method of sample size determination, a substantial amount of cost savings can be expected by using the suggested sample sizes.

Practical implications – The paper presents a structured approach of opportunity assessment for improving service quality from a strategic alignment perspective, particularly in the five dimensions: tangibles, reliability, responsiveness, assurance, and empathy. The new approach provides practising managers with a decision-making tool for measuring service quality, detecting problematic situations and selecting the most urgent improvement project. Once the existing service problems are identified and improvement projects are prioritized, it can lead to the direction of continuous improvement for any service industry.

Originality/value – Given a managerial target on any desired service level as well as customers’ perceptions and expectations, the new KPI could be applied to any non-normal service quality and other survey data. Thus, the corporate performance in terms of key factors of business success can also be measured by the new KPI, which may lead to managing complexities and enhancing sustainability in service industries.

Keywords Process analysis, Service industries, Service quality assurance

Paper type Research paper

1. Introduction

In recent years, the value produced by service industries has become a main source of social wealth, and has exceeded the corresponding wealth generated by manufacturing industries in most advanced countries. Job opportunities and the creation of wealth in the service sector show continuous growth in these countries. Moreover, the concept that the quality of interaction while procuring a product or service has a direct impact on...
customer satisfaction is now prevailing in marketing theory. This concept puts all business activities in the arena of service quality since business activity is the exchange of intangible goods as well as, but not exclusively, tangible goods regardless of the product being sold. Hence, it is necessary to develop a quantitative method for measuring the service quality of customer perceptions through a key performance index or indicator (KPI). Using the highly successful method of product quality assurance, we propose a new KPI and its interval estimation for measuring service quality.

A significant amount of research has been carried out since service quality concepts were introduced by Grönroos (1982). In most surveys done on service quality, the difference between customer expectation and perception is called GAP5 based on PZB model (Parasuraman et al., 1985). Since then, service quality revolves around their idea and it is defined as the result of the comparison that customers make between their expectations about a service and their perception of the way the service has been performed. Researchers generally use a t-test or the Mann-Whitney test to determine whether GAP5 is zero or not. In view of the unique characteristics of service quality, we combine this concept with a non-normal unilateral index from the manufacturing industry to evaluate GAP5. We also propose that the performance of various service industries can be evaluated by such an index.

In this paper, we first examine the evolution of Process Capability Indices used in manufacturing to design a KPI for the service industry. Then, we employ a bootstrap method to construct the performance confidence interval. The coverage probability is used to evaluate the index reliability. From the interval estimation, managers using our KPI can immediately evaluate the performance of their service. Also the sample size required for measuring service quality can be determined. Finally, we use a realistic example from a service industry in Taiwan to evaluate the five dimensions of the SERVQUAL model (Parasuraman et al., 1988). In GAP5, the difference between customer service perception and his/her service expectation is analyzed. Since data may range from normal to non-normal distributions in each of the five dimensions, the estimated KPI and its confidence limits are calculated based on their underlying distribution. We further demonstrate that service quality problems can be detected easily by using the new KPI and appropriate diagrammatic means.

2. Literature review
2.1 Evolution of process capability indices
Process capability indices have been used widely in industries to provide quantitative measures on process performance. The four main normal process capability indices used with normal distributions are defined below. For a bilateral specification:

\[
C_p = \frac{USL - LSL}{6\sigma}
\]

\[
C_{pk} = \min\left(\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right)
\]

\[
C_{pm} = \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}}
\]
\[ C_{pmk} = \min \left( \frac{USL - \mu}{3\sqrt{\sigma^2 + (\mu - T)^2}} \cdot \frac{\mu - LSL}{3\sqrt{\sigma^2 + (\mu - T)^2}} \right) \]

Measuring service quality

For a unilateral specification:

\[ C_p = C_{pk} = C_{pl} = \frac{\mu - LSL}{3\sigma} \quad \text{(the smaller the better)} \]

\[ C_p = C_{pk} = C_{pl} = \frac{\mu - LSL}{3\sigma} \quad \text{(the larger the better)} \]

Here \( LSL \) is the lower specification limit, \( USL \) is the upper specification limit, \( T \) is the target value, \( \mu \) is the process average, and \( \sigma \) is the process standard deviation. The \( C_p \) index considers the ratio of the allowable tolerance to its natural variation (Juran, 1974), and thus only reflects process precision. The \( C_{pk} \) index not only considers the process precision, but also its accuracy (Kane, 1986). However, the \( C_p \) and \( C_{pk} \) indices are independent of the target \( T \), which may fail to provide information on its variation. The \( C_{pm} \) index was developed by taking into account the deviation of the process mean from its target value. He then revised his capability index to \( C_{pmk} \) to deal with asymmetric tolerance shifts between the target and two specification limits.

Under the assumption that process data follows a normal distribution, the indices above work well when evaluating the capability of a manufacturing process. However, numerous samples exhibit non-normal distributions. Gunter (1982) pointed out that defects falling outside \( \pm 3\sigma \) specification limits are significantly different in the case of three particular non-normal distributions (i.e. chi-square distribution, \( t \) distribution and uniform distribution) even with the same \( C_p \) and \( C_{pk} \) indices. Clement (1989) used a Pearson distribution curve to estimate a non-normal process capability index. If the distribution of the measurements of a quality characteristic belongs to this Pearson family of probability curves, which consists of normal, lognormal, \( t \), \( F \), beta and gamma distributions, then the four common non-normal process capability indices can be defined as follows. For a bilateral specification:

\[ C_p = \frac{USL - LSL}{\xi_{99.865} - \xi_{0.135}} \]

\[ C_{pk} = \min \left( \frac{USL - M}{\xi_{99.865} - M} \cdot \frac{M - LSL}{\xi_{0.135}} \right) \]

Here \( \xi_{99.865} \) and \( \xi_{0.135} \) are the 99.865 and 0.135 percentiles of the Pearson family, \( M \) is the process median. Pearn and Chen (1995) further defined the non-normal \( C_{pm} \) and \( C_{pmk} \) as:

\[ C_{pm} = \frac{USL - LSL}{6\sqrt{(\xi_{99.865} - \xi_{0.135}/6)^2 + (M - T)^2}} \]
\[ C_{pm} = \frac{\min(USL - M, M - LSL)}{3\sqrt{\left(\frac{\xi_{99.865}}{\xi_{0.135}}/6\right)^2 + (M - T)^2}} \]

The \( C_{pm} \) uses the tolerance of Six Sigma, \( \pm 3\sigma \), in the denominator, while \( C_{pmk} \) uses 3 Sigma since it takes the minimum deviation of the numerator into consideration. For a unilateral specification:

\[ C_p = C_{ph} = C_{pu} = \frac{USL - M}{\xi_{99.865} - M} \quad (\text{the smaller the better}) \]

\[ C_p^l = C_{pl} = \frac{M - LSL}{M - \xi_{0.135}} \quad (\text{the larger the better}) \]

2.2 Evolution of bootstrap confidence intervals

Efron (1979) introduced and developed a non-parametric, computer intensive estimation method called bootstrapping. It is a data-based simulation method for statistical inference, which can be used to produce inferences to the true value of \( \theta \), lies in the interested interval with a given confidence level. The method, illustrated in the second step in Figure 1, is described as follows. Suppose \( (x_1, x_2, \ldots, x_n) \) is an original

---

**Figure 1.**
Procedure for calculating the bootstrapping interval and coverage probability

1. Original data \( x_1, x_2, \ldots, x_n \) are given
2. Draw each \( x_i^*, i = 1, 2, \ldots, n \) with replacement from the original data points \( x_1, x_2, \ldots, x_n \)
3. Estimate true KPI, \( \hat{\theta}_b \)
4. Construct confidence limits of KPI from \( \hat{\theta}_b^*, b = 1, 2, \ldots, B \) using Biased-Corrected Acceleration Percentile Bootstrap
5. If \( m \) of 1,000 confidence limits covers the real KPI, the coverage probability is \( \frac{m}{1,000} \times 100\% \)
sample with size \( n \), and \( \theta \) is the parameter that we are interested in (i.e. the KPI). A bootstrap sample, denoted by \((x_1, x_2, \ldots, x_n)\) is a sample of size \( n \) drawn with replacement from the original sample, i.e. Draw each \( x_i^* \), \( i = 1, 2, \ldots, n \) with replacement from the original data points \((x_1, x_2, \ldots, x_n)\). Repeat this procedure \( B \) times, we can obtain \( B \) bootstrap samples and \( B \) bootstrap estimates, denoted as \( \hat{\theta}_b, b = 1, 2, \ldots, B \). There are four types of bootstrap confidence intervals (Efron and Tibshirano, 1993), namely, the standard bootstrap confidence interval (SB), the percentile bootstrap confidence interval (PB), the biased-corrected percentile bootstrap confidence interval (BCPB), and the biased-corrected acceleration percentile bootstrap confidence interval (BCa). The BCa method developed by Efron and Tibshirano (1993) adjusts the PB (percentile bootstrap given the approximated distribution) confidence interval by the bias-correction and the acceleration, where

\[
\hat{\theta}_0 = \Phi^{-1} \left( \sum_{b=1}^B I(\hat{\theta}_b - \bar{\theta})/B \right); \quad \hat{\theta}_b, b = 1, 2, \ldots, B \text{ are bootstrap estimators; } \bar{\theta} \text{ is an estimator of KPI from original data; } \Phi \text{ is the standard normal cumulative density function; i.e. } z_\alpha = \Phi^{-1}(\alpha).
\]

\[
\text{Acceleration } \hat{a} = \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^2/6 \{ \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^2 \}^{3/2}, \quad \hat{\theta}_i \text{ is an estimator of KPI from } x_i \text{ deleted; } \bar{\theta} = \sum_{i=1}^n \hat{\theta}_i/n.
\]

Since BCa is the most commonly used bootstrapping method, the bootstrap confidence interval with a biased-corrected acceleration percentile (BCa) is adopted in this paper. On the other hand, coverage probability of a confidence interval is the proportion of the time that the interval contains the true value of interest. It is equivalent to the confidence level of the constructed interval, which is effectively the “nominal coverage probability” of the procedure for constructing confidence intervals. The “nominal coverage probability” is often set at 0.95. In other words, the coverage probability is the actual probability that the interval contains the true value of interest. It is equivalent to the confidence level of the constructed interval, which is effectively the “nominal coverage probability” of the procedure for constructing confidence intervals. The “nominal coverage probability” is then described in Figure 1.

When the process data follow normal distribution, the interval estimation for \( C_p, C_{pk}, C_{pm} \) and \( C_{pmk} \) indices are constructed by taking sampling error into account. However, when processes data follow non-normal distribution, the sampling distributions of PCI’s estimators and their confidence interval become very difficult to obtain. Franklin and Wasserman (1992) showed that the behavior of the first three bootstrap confidence intervals for \( C_p, C_{pk} \) and \( C_{pm} \) indices are reliable whether the underlying process distributions were normal, skewed, or heavy tailed. Price and Price (1993) then estimated the confidence limits of \( C_{pk} \) using the adjusted bootstrap method. Tong and Chen (1998) also employed the SB, PB, and BCPB methods to construct lower confidence limits for the \( C_p, C_{pk} \) and \( C_{pm} \) indices under non-normal distributions. Choi et al. (1996) also devised an estimator and an asymptotic distribution by the bootstrapping method. Chen and Tong (2003) and Chen and Chen (2004) constructed confidence limits for \( C_{pk_1} - C_{pk_2} \) and \( C_{pm_1}/C_{pm_2} \) to compare the difference between two processes. Puga-Leal and Pereira (2007) assumed service performance follows a normal distribution and proposed a new service capability index. Due to the fact that the service quality data may not follow normal distribution, a KPI under non-normal distribution is considered in this paper.

### 2.3 Evolution of service quality analysis: PZB and the SERVQUAL model

Service has three characteristics: intangibility, heterogeneity, and inseparability. When purchasing goods, the consumer employs many tangible cues to judge quality: style, hardness, color, label, package, and fit. When purchasing services, however, fewer
tangible cues exist. In most cases, tangible evidence is limited to the service provider’s physical facilities, equipment, and personnel (Parasuraman et al., 1985).

The PZB model was developed by Parasuraman et al. in 1985. This model consists of five gaps where enterprises fail in meeting both internal and external expectations. GAP1 to GAP4 reflect the service marketer’s internal patterns and potential deficiencies, while GAP5, which will be our focus, represents the gap perceived by the customer, the difference between his expectation and perception of a given service.

Although there are various types of service quality tests, the SERVQUAL instrument, also developed by Parasuraman, is widely used in the service industry. Parasuraman’s research reveals that a consumer’s assessment of service quality fits into ten potentially overlapping dimensions. These dimensions are tangibles, reliability, responsiveness, communication, credibility, security, competence, courtesy, understanding or knowing the customer, and accessibility. To simplify these ten dimensions, Parasuraman et al. (1988) removed the items with relatively low item-to-total correlations and developed a refined scale with 22 items spread over the following five dimensions:

1. **Tangibles.** Physical facilities, equipment, and appearance of personnel.
2. **Reliability.** Ability to perform the promised service dependably and accurately.
3. **Responsiveness.** Willingness to help customers and provide prompt service.
4. **Assurance.** Knowledge and courtesy of employees and their ability to inspire trust and confidence.
5. **Empathy.** Caring, the individualized attention the firm provides its customers.

Since the five gaps are also important to our analysis, a description is given in Table I and the key insights gained from surveys, focus groups and executive interviews regarding the GAP effect on service quality. Taylor and Cronin (1994) attempted to offer an alternative of measuring service quality, call SERVPERF. The development of

<table>
<thead>
<tr>
<th>Gap</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer expectation-management perception gap (GAP1)</td>
<td>The gap between consumer expectations and management perceptions of those expectations will have an impact on the consumer’s evaluation of service quality</td>
</tr>
<tr>
<td>Management perception-service quality specification gap (GAP2)</td>
<td>The gap between management perceptions of consumer expectations and the firm’s service quality specifications will affect service quality from the consumer’s viewpoint</td>
</tr>
<tr>
<td>Service quality specification-service delivery gap (GAP3)</td>
<td>The gap between service quality specifications and actual service delivery will affect service quality from the consumer’s standpoint</td>
</tr>
<tr>
<td>Service delivery-external communications gap (GAP4)</td>
<td>The gap between customer feedback and action within the service delivery departments generates the perception of acknowledgement</td>
</tr>
<tr>
<td>Expected – perceived service gap (GAP5)</td>
<td>Focus groups unambiguously supported the notion that the key to ensuring good service quality is meeting or exceeding consumers’ expectation from the service</td>
</tr>
</tbody>
</table>

Table I. Description of the five gaps
SERVPERF model aimed to provide a method of measuring perceived service quality and the significance between service quality, customer satisfaction and purchase intentions. Both SERVQUAL and SERVPERF instruments have been widely applied in miscellaneous disciplines, such as information technology, health care, education, and transportation, etc. dos Santos et al. (2009) applied the SERVQUAL instrument to several customers of software providers and appraised in a CMM/CMMI (Capability Maturity Model/ Capability Maturity Model Integration). Their results showed a considerable discrepancy between customers’ expectations and their perceptions of the services provided. Roses et al. (2009) evaluated the perception gaps of service quality between information technology (IT) service providers and their clients by using both SERVPERF of SERVQUAL instruments. Their study added some enhancements to the SERVQUAL model. Lin et al. (2009) showed that the SERVQUAL instrument is a useful measurement tool in assessing and monitoring service quality in health care, which enables medical staffs to identify the most needed improvement areas from the patients’ perspectives. Chatterjee et al. (2009) adopted a SERVQUAL-type approach by using students’ feedback report as a valid measure of teaching effectiveness to improve the teaching quality. Rosenbaum and Wong (2009) combined the product-focused return-on-marketing framework with SERVQUAL instrument to study consumers’ future behavioral intentions under the moderating influence of ethnocentrism. Nadiri et al. (2009) proposed a conceptual model called “BANKZOT” (a modified version of SERVQUAL) to measure the tolerance zones in the banking sector. Their results revealed that the two larger gaps in the five dimensions of service quality are tangibles and empathy provided by the banks. Fornell (1992) developed a customer satisfaction barometer (CSB) and used it to measure the actual levels of service quality across 30 industries and 100 companies. However, according to Gilmore and McMullan (2009), none of the above-mentioned models have enjoyed the same degree of use and adaptation as SERVQUAL. Thus, the “perceptions-minus-expectations” measurement of SERVQUAL is adopted in this research.

3. Design methodology

3.1 Development of the new key performance index (KPI)

As can be seen from Table I, GAP5 depends on customers’ expectations and perceptions. Since the distance between expectations and perceptions will determine customer’s perception of service quality, we expect a positive response to the service quality when customers’ perception of service is higher than their expectation. On the other hand, if customers’ perception of service is lower than their expectation, we would expect a negative response to the service quality.

For our purposes, suppose that \( A \) = customers’ expected adequate service, \( P \) = customers’ perceptions, and \( GAP = P - A \). Since \( GAP \) is considered to be a relative measure of service quality, the larger the \( GAP \) means the better service quality will be, i.e. the customers’ perception will outstrip the expectation. But if the expectation is higher than the perceived performance, then this \( GAP \) will be negative.

Since \( GAP \) may not follow normal distribution, we re-introduce the non-normal process capability index proposed by Clement (1989), which uses percentile ranges to establish \( C_p \) and \( C_{pk} \). In the manufacturing case, it is common to set the percentile \( q = 0.135 \) and \( 1 - q = 99.865 \), which gives a percentile range equivalent to 6\( \sigma \), the six sigma standard. The range can be appropriately adjusted given the sample data.
while $d$, the semi-distance between the upper and lower specification limits, is also adjustable and depends on the specification limits. Thus, equations (1) and (2) can be expressed as follows:

$$C'_p = \frac{USL - LSL}{\xi_{100-q} - \xi_q} = \frac{2d}{\xi_{100-q} - \xi_q}$$  \hspace{1cm} (3)

$$C'_{pk} = \min\left(C'_{pu}, C'_{pl}\right) = \min\left(\frac{USL - M}{\xi_{100-q} - M}, \frac{M - LSL}{M - \xi_q}\right)$$  \hspace{1cm} (4)

Combining Parasuraman’s concept with Clement’s non-normal process capability index in equation (4) where the GAP is “the larger the better,” we propose a key performance index based on the $C_{pk}$ lower specification limit, LSL, above. It is defined as follows:

$$KPI = \frac{M_{\text{GAP}} - L}{M_{\text{GAP}} - \xi_{\text{GAP}}(q)}$$

where $M_{\text{GAP}}$ is the median of GAP, $\xi_{\text{GAP}}(q)$ is the $q$th percentile of GAP, and $L$ is our lower specification limit or the minimum desired service level. In other words, we expect that the GAP should not be smaller than $L$ when evaluating the service performance of an enterprise. Note that $L = 0$ indicates the customer’s minimum expectations were met. The value of $100 - q$ will depend on an enterprise’s target service level or its service objective. For instance, when the service objective is set such that the perception of 95 percent of customers (i.e. $(100 - q)$ percent is 0.95) will be greater than the minimum desired service level $L$, then $q$ is set to 0.05.

When $KPI = 1$, then $\xi_{\text{GAP}}(q) = L$, the service performance achieves the minimum desired service level. Hence, if $(100-q)$ percent of the GAP is greater than $L$, then $KPI$ is greater than 1. Conversely, when the $KPI$ is lower than 1 the enterprise does not meet its minimum desired service level, which is consistent with the process capability index used in manufacturing.

3.2 Evaluating the reliability of the bootstrap confidence level of KPI

To construct the confidence interval and coverage probability for our KPI, we use the BCa method because it is considered to be the most correct one in bootstrapping. The procedure for calculating the bootstrap confidence interval and coverage probability is shown in Figure 1.

3.3 Procedure for calculating the bootstrap confidence intervals

For testing non-normal distribution of data, we use a Beta distribution with five possible conditions in order to preserve the generality of our result. These conditions are described as follows:

1. Extremely skewed to the right: Beta (4, 10).
2. Skewed to the right: Beta (3, 4).
4. Skewed to the left: Beta (4, 3).
5. Extremely skewed to the left: Beta (10, 4).
With $\alpha = 0.05$, a standard 95 percent statistical confidence level, the frequency of coverage takes the form of a binomial random variable with $p = 0.95$ and $n = 1,000$. Thus, to achieve a 99 percent confidence interval in our coverage proportion means that the tolerance range is $0.95 \pm z_{0.005} \sqrt{(0.95)(0.05)/1,000} = 0.95 \pm 0.0178$. Hence, we can be 99 percent confident that a “true 95% confidence limit” proportion of coverage will fall between 0.933 and 0.967. Figures 2-4 indicate the coverage probabilities of BCa confidence limits for five beta distributions with different sample sizes and service levels (90 percent, 95 percent, and 99 percent, respectively) at the 95 percent confidence level.

3.4 Suggested sample size under different confidence and service levels
Sample size affects the cost of quality control in both in the manufacturing and the service industries. Of course statistically, the larger the sample size, the more accurate an estimate will be, too large a sample usually results in excessive cost and wasted materials. Hence, determination of sample size is a critical issue. We used the following procedures to determine sample size for reliable KPIs:

- Decide on the desired service level.
- Calculate coverage probabilities for five Beta distributions using various sizes and establish an average coverage probability line.
- Examine the trend of the average coverage probability line and determine the correct sample size, as long as the trend line reaches a steady state and the average coverage probability of the five Beta distributions starts to fall within the required confidence interval.

![Figure 2](image-url)
Figure 3. Coverage probabilities for five Beta distributions with different sample sizes at the 95 percent target service level.

Figure 4. Coverage probabilities for five Beta distributions with different sample sizes at the 99 percent target service level.
For instance, with a desired 95 percent service level at $\alpha = 0.05$, we consider the above criteria that the trend line reach a steady state and the average coverage probability of the five Beta distributions starts to fall within the 95 percent confidence interval. Then, 250 is the correct sample size for a 95 percent target service level at 95 percent confident level. We constructed 25 Beta distributions either skewed to the right, skewed to the left, or symmetric. Then, using the above criteria, the correct sample size was determined for various combinations of confidence and target service levels. These simulation results are summarized in Table II. We conclude that with higher target service levels, larger sample sizes are required for a 95 percent coverage probability.

To reach a steady state, as shown in Figures 2-4, 200 samples are required at the 90 percent target service level, whereas 250 samples are required at the 95 percent target service level, and so on. Moreover, with a 99 percent target service level at $\alpha = 0.05$, the average coverage probability line reaches the steady state and its lower confident limit equals 0.933 (note that at 95 percent confident level, the 95 percent coverage probability falls between 0.933 and 0.967) when sample size is 450. Hence, BCa confidence intervals can meet a 95 percent coverage probability under the 99 percent target service levels. Likewise, with the 99 percent target service level set at $\alpha = 0.01$, where the average coverage probability line reaches the steady state and its lower confident limit equals 0.982 when sample size is 650. Since a large number of survey respondents are normal and likely in the service industry, we consider the relatively large sample sizes above to be realistic, and also necessary for valid KPIs.

Traditionally, the sample size determination formulas come from the formulas listed below for the maximum error ($e$) of the estimates. The formula is solved for $n$.

Given $(1 - \alpha)$ 100 percent confidence interval of population proportion $\hat{p}$, $\hat{p} \pm Z_{\alpha/2}\sqrt{\hat{p}(1 - \hat{p})/n}$; then the maximum error $e = Z_{\alpha/2}\sqrt{0.5(1 - 0.5)/n}$.

Hence $n = (Z_{\alpha/2} \times 0.5/e^2)^2$.

Since the sample has not been taken, there is no value for the sample proportion. If there is no previous study or estimate available, then we use 0.5 for $\hat{p}$. Based on the fact that $\hat{p}$ follows a normal distribution, the required sample size under different combinations of service and confidence levels can be calculated accordingly. The required sample sizes under different combinations of confidence and target service levels for conducting a survey using traditional method is then summarized in Table III.

Suppose that the sampling cost per questionnaire is $10 and the equivalent confidence level (comparable to the coverage probability) is 90 percent, then a maximum cost savings of $6520 can be expected by comparing the sample sizes listed in Tables II and III. Similarly, a maximum cost savings of $8680 can be expected if the equivalent confidence level is 95 percent and a maximum cost savings of $16,440 can be expected if the equivalent confidence level is 99 percent.

### Table II.
The suggested sample sizes under different combinations of coverage probabilities and target service levels using bootstrapping method.
4. Managerial analysis and discussion

4.1 Developing service quality assurance

The purpose of service quality assurance is to assure that the service quality will satisfy the requirements of target customers. To satisfy the requirements of target customers in service industries, the suppliers need create value to the customers and have a deep understanding of the customers as well as the problematic situations on hand. It is of importance the demands of the customers are defined in a correct way. The KPI and its estimated intervals can be served as a useful tool to measure the gap between the existing supplier’s performance and the customer’s requirements/expectations. Once the problematic situations have been identified and quantified, the supplier usually form a quality improvement team and apply Six Sigma’s DMAIC (Define, Measure, Analyze, Improve, Control) methodology to solve the identified problem. Hence, the service quality can be improved through the Six Sigma project. For example, if the queuing time and maintenance time have been identified as the KPI of the internal processes of a service industry, then we can perform a process capability study to calculate/compare their capability indices before and after quality improvement. Moreover, given a managerial target on any desired service level as well as customers’ perceptions and expectations, the new KPI could be applied to any non-normal service quality and other survey data. Thus, the corporate performance in terms of the key factors of business success, such as product and service quality, productivity and operation efficiency, research capability and innovation, corporate social responsibility, customer satisfaction, etc. can also be measured by the new KPI, which may lead to managing complexities and enhancing sustainability in service industries. Listed below are the procedures for evaluating the performance of service quality. It provides a useful reference for practicing managers in identifying the problematic situations and focusing their improvement efforts:

1. Design a sampling plan based on Table II and conduct a questionnaire survey.
2. Perform statistical data analysis and testing the goodness of fit for five dimension distributions of service quality.
3. Calculate the KPIs for the five dimensions of service quality, i.e. Tangibles, Responsiveness, Assurance, Reliability, Empathy.
4. Establish the Bootstrap confidence intervals for KPIs based on the procedures shown in Figure 1.
5. Evaluate the adequacy of KPIs by using the above confidence intervals or finding the corresponding actual service level from Table IV as a diagnostic mean.

Once the problematic situations are detected, one can employ various Lean Six Sigma techniques to conquer complexities and achieve cost reductions/customer satisfaction.

<table>
<thead>
<tr>
<th>Target service level</th>
<th>Confidence level (1 – α)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.9</td>
</tr>
<tr>
<td>90%</td>
<td>752</td>
</tr>
<tr>
<td>95%</td>
<td>752</td>
</tr>
<tr>
<td>99%</td>
<td>752</td>
</tr>
</tbody>
</table>

Table III.
The required sample sizes under different combinations of confidence and target service levels using traditional method.
in service industries. The following realistic example is given to illustrate how the
service performance can be evaluated by using this structured approach of service
quality assurance.

4.2 A realistic example
In order to demonstrate and verify our calculation of KPI, we use a realistic data set
provided by Lian (2008). Based on SERVQUAL instrument, he developed the survey
questionnaires including important service items for the consumers, their actual
cognition of service quality provided, the customer’s revisit intention, and an indication
of the gap between business owners’ perception of service provided and the consumers.
He received a total of 164 usable questionnaires (according to Table II, 150 samples is
required at 95 percent target service level and 90 percent coverage probability) with 22
items distributed across the five SERVQUAL dimensions: tangible, responsiveness,
assurance, reliability, and empathy. He used a seven-point Likert Scale. After collecting
the survey questionnaires, a factor analysis was conducted using the difference
between the re-scaled factor loadings as the GAP. Then, percentile grouping were
made and KPI’s were calculated/ determined. We test the data on each of the five
dimensions to find its Beta distribution by using a parametric sweep of the shape and
scale parameters from 1 to 20. Using the Kolmogorov-Smirnov test (K.S. test) on each
combination of shape and scale, we select the distribution, which has the largest p
value as the most appropriate distribution for each dimension. The results of the
Kolmogorov-Smirnov test for each dimension are presented in Table V. Table IV
summarizes the corresponding actual service levels for various KPIs under different
combinations of Beta (10, 15), Beta (11, 16) and target service levels, (100 - q)%.

<table>
<thead>
<tr>
<th>(100 - q)%</th>
<th>Beta (10, 15)</th>
<th>Actual service level</th>
<th>(100 - q)%</th>
<th>Beta (11, 16)</th>
<th>Actual service level</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>-0.67</td>
<td>0.211</td>
<td>90%</td>
<td>-0.67</td>
<td>0.210</td>
</tr>
<tr>
<td>0.00</td>
<td>0.500</td>
<td></td>
<td>0.00</td>
<td>0.500</td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>0.799</td>
<td></td>
<td>0.67</td>
<td>0.800</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>0.900</td>
<td></td>
<td>1.00</td>
<td>0.900</td>
<td></td>
</tr>
<tr>
<td>1.33</td>
<td>0.960</td>
<td></td>
<td>1.33</td>
<td>0.960</td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>-0.67</td>
<td>0.157</td>
<td>95%</td>
<td>-0.67</td>
<td>0.155</td>
</tr>
<tr>
<td>0.00</td>
<td>0.500</td>
<td></td>
<td>0.00</td>
<td>0.500</td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>0.857</td>
<td></td>
<td>0.67</td>
<td>0.858</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>0.950</td>
<td></td>
<td>1.00</td>
<td>0.950</td>
<td></td>
</tr>
<tr>
<td>1.33</td>
<td>0.989</td>
<td></td>
<td>1.33</td>
<td>0.988</td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>-0.67</td>
<td>0.088</td>
<td>99%</td>
<td>-0.67</td>
<td>0.085</td>
</tr>
<tr>
<td>0.00</td>
<td>0.500</td>
<td></td>
<td>0.00</td>
<td>0.500</td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>0.930</td>
<td></td>
<td>0.67</td>
<td>0.931</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>0.990</td>
<td></td>
<td>1.00</td>
<td>0.990</td>
<td></td>
</tr>
<tr>
<td>1.33</td>
<td>1.000</td>
<td></td>
<td>1.33</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

Table IV. Summary of the corresponding actual service levels for various KPIs under different combinations of Beta (10, 15), Beta (11, 16) and target service levels, (100 - q)%.
use $L = 0$ as the minimum desired service level of the GAP indicating that, as a baseline, no difference between customers’ perceived service and adequate service existed, i.e. the customer’s minimum expectations were met. The confidence limits then can be used for two purposes. First, they can be used to test the following hypothesis:

$H_0$. KPI = 1

$H_1$. KPI $\neq$ 1

If the range of the confidence interval does not include 1, we can reject the null hypothesis. In other words, the KPI estimator will be significantly greater than or lower than 1. Hence, we can determine whether a service dimension is greater than or lower than its desired level by observing whether the upper or lower limits of the confidence interval is appropriately placed. Secondly, a manager can observe exactly where within this interval the actual service level resides.

In Figure 5, we find that only the confidence interval of KPI of the Tangible dimension is clearly greater than 1, under a 90 percent target service level, which

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Distribution</th>
<th>$p$ value from KS test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tangibility</td>
<td>beta $(10, 15) \times 10 - 4$</td>
<td>0.957</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>beta $(11, 16) \times 10 - 4$</td>
<td>0.413</td>
</tr>
<tr>
<td>Assurance</td>
<td>beta $(11, 16) \times 10 - 4$</td>
<td>0.277</td>
</tr>
<tr>
<td>Reliability</td>
<td>beta $(11, 16) \times 10 - 4$</td>
<td>0.348</td>
</tr>
<tr>
<td>Empathy</td>
<td>beta $(11, 16) \times 10 - 4$</td>
<td>0.347</td>
</tr>
</tbody>
</table>

Table V. The Kolmogorov-Smirnov test results for five dimensions of service quality

Notes: The distribution was further scaled for percentile ranging by multiplying by ten and subtracting four as can be seen above.

Figure 5. Estimated KPI and its confidence limit for the five dimensions under a 90 percent target service level.
indicates the Tangible dimension is significantly meeting the requirement. In Table IV, given the Tangible KPI = 1.112, one can obtain the actual service level is 92 percent by interpolation. By contrast, the KPI confidence interval of the four dimensions (Responsiveness, Assurance, Reliability, Empathy) are all significantly less than 1 indicating that they do not meet the 90 percent target service level. Similarly, in Table IV, one can obtain the actual service levels for Responsiveness, Assurance, Reliability, and Empathy are equal to 81.7 percent, 82.5 percent, 77.6 percent and 79.76 percent respectively by interpolation. In Figure 6, we find that the KPI confidence interval for the Tangible dimension includes 1, under a 95 percent target service level. In Table IV, given the Tangible KPI = 0.9291, one can obtain the actual service level is 93 percent by interpolation. By contrast, the KPI confidence limits of the other four dimensions are all significantly less than 1, indicating that they do not meet the 95 percent target service level. Similarly, in Table IV, the actual service levels for Responsiveness, Assurance, Reliability, and Empathy are equal to 79.7 percent, 80.97 percent, 77.3 percent and 77 percent respectively by interpolation. This means the Responsiveness, Assurance, Reliability and Empathy of service quality need to be further improved.

5. Conclusions and managerial implications
Service quality is a matter of finding out what creates value to customer and then offering it. This requires familiarity with the customer and deep understanding of the problematic situation on hand. In this paper, we show that KPI plays an important role in service quality assurance since it provides a quantitative measure of service quality. For measuring service quality and enhancing customers’ satisfaction, we have developed a new key performance index based on the capability indices which are widely used in manufacturing for process control. In particular, we used Clement’s non-normal process capability index with a unilateral specification. In addition, we
applied Efron’s bootstrapping method to determine sample sizes required for conducting process capability study and/or questionnaire survey. Moreover, compared to the traditional method of sample size determination, we have shown that a substantial amount of cost savings can be expected by using our suggested sample sizes. In the realistic example, the new KPI is applied to SERVQUAL data using Parasuraman’s five dimensions to determine its managerial efficacy. We used $L = 0$, i.e. the customer’s minimum expectations were met, to demonstrate the variance of actual service performance levels (in five SERVQUAL categories) given 90 percent and 95 percent target service levels. Our new KPI performed well in this regard. Also, by using the new KPI and its confidence intervals, one can discern if the desired service level has been achieved given that tolerance zone in which service is actually being delivered. Hence, this new approach allows managers to evaluate both the relative importance values of existing KPIs and their achievability levels for the sake of setting priorities during the implementation of the possible improvement projects. Once the existing service problems are identified and improvement projects are prioritized, it can lead to the direction of continuous improvement for any service industry. Therefore, we conclude that since the capability index approach has been very successful in the manufacturing industry, its application to the ever-expanding service industries is a step towards quantifying the dimensions of service quality. In overall, this research presents a structured approach of opportunity assessment for improving service quality from a strategic alignment perspective, particularly in the five dimensions: tangibles, reliability, responsiveness, assurance, and empathy. It not only can be extended to almost any service industry that surveys its clientele, but also provides practicing managers a useful decision-making tool for measuring service quality, detecting problematic situations and selecting the most urgent improvement project.

References


Further reading

About the authors
Jeh-Nan Pan is currently a Director and Vice Dean at College of Management, National Cheng-Kung University. His expertise is in the fields of quality management and reliability engineering, operations research and decision science, industrial statistics and statistical consulting. He has founded the first Statistical Consulting Center in Taiwan and has served as a chairman of Statistics Department at the college of Management, NCKU. He was previously on the staff of USC and California State University in the USA and has more than ten years of industrial experience, starting as a Senior Quality Assurance Engineer and then promoted to Chief Statistician at Leach Corporation and Statistical Process Control Program Manager at Teledyne Inc. in Los Angeles, USA. He has authored two textbooks: Quality Management: Principles and Practice and Proactive Quality Assurance (in Chinese) and coauthored (with Dr William Kolarik) a textbook Creating Quality (McGraw Hill Inc.) He has published over 70 papers both in domestic and international journals such as: International Journal of Production Research, IEEE Transaction on Reliability, European Journal of Operational Research, Computers and Industrial Engineering, Quality Engineering, International Journal of Quality and Reliability Management, Industrial Engineering Management and Data System, Quality and Reliability Engineering International, Microelectronic and Reliability, etc. Currently, he serves on the editorial boards of Journal of Quality, International Journal of Probability and Statistical Science, Management and System Journal, APMR and IJMED. In addition to pursuing academic research, he is also very active in offering short courses and consulting with industry. He holds M.S.-I.E. and PhD degrees from Texas Tech University in Lubbock, Texas. He is an ASQC-certified Quality and Reliability Engineer. He has received several distinguished awards including Research and Teaching Excellence awards from NCKU, four best paper awards from CSQ, etc. Jeh-Nan Pan is the corresponding author and can be contacted at: jnpan@mail.ncku.edu.tw

Tzu-Chun Kuo is a graduate student at National Cheng Kung University, Tainan, Taiwan. Her chosen field is statistics.

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com
Or visit our web site for further details: www.emeraldinsight.com/reprints