A Multitrait Multimethod Analysis of Service Quality Measurement among Users of Cloud-Based Service Platforms

**Abstract**

This research furthers our understanding of whether consumers of cloud-based service platforms can distinguish between gap theory dimensions (i.e., expectation and performance) of these platforms. We build upon and extend the work of (Natesan and Aerts 2016) by applying confirmatory factor analysis on gap scores from survey data to develop and test an improved approach of measuring service system quality in cloud-based service platforms. Using the IS-adapted SERVQUAL instrument, we apply the correlated uniqueness model, which is part of the multitrait multimethod (MTMM) framework, to evaluate the validity of using GAP scores and account for methods effect. There is significant support for method effects as shown by our suggested model paths in the trait convergent validity model with medium-large factor coefficients. Additionally, the fit of correlated uniqueness model indicate respondents can distinguish between the gap theory dimensions of the IS-adapted SERVQUAL instrument. Using our measurement approach by incorporating error correlations could make a meaningful impact service quality measurement practice.

**Keywords:** information system service quality, difference scores, gap measures, multitrait multimethod models, IS-adapted SERVQUAL

**INTRODUCTION**

Information systems, retail and consumer service research have experienced an increase in use and relative importance of empirical data to help develop and test theories. However, it is important to note that much of these empirical works are anchored in survey-based methodologies where data is collected in the form of scale items and analyzed to measure a latent variable (Malhotra and Sharma 2008). For example, information systems, operations, and marketing managers use survey data to gather information from customers to analyze and determine functional areas within service operations that do not add value to the customer service experience (Boakye et al., 2014; Grover et al. 1996). As a result, either preexisting scales that have already been developed or new ones developed are used to measure the extent to which customers assess and evaluate their satisfaction and dissatisfaction with their service experience. Consistent with such evaluation is the use of the IS-adapted SERVQUAL instrument to evaluate the difference between IS service quality performance and expectation (Jiang, Klein, Ron, and Lin, 2001; Rosene, 2003; Gorla, 2011).

While service quality is an overall attitude exhibited by a service firm and a critical construct towards customer (internal and external) satisfaction, loyalty, firm performance, among others, the ability to conceptualize and measure it has been elusive (Parasuraman et al., 1985). It is in light of this difficulty that SERVQUAL’s gap measure was developed. SERVQUAL is a survey instrument originally developed by marketing researchers to assess the service quality based on the gap between expected service and perceived service delivery. The SERVQUAL consists of 22 items, forming 5 major dimensions used by customers to evaluate the quality of service they experience at the hands of the service provider (Parasuraman et al., 1985). It is also important to note that SERVQUAL has also been used extensively in operational contexts as well as marketing (Chase and Apte, 2007; Zeithaml et al., 1990).

In the last decade, there have been a considerable number of debates, arguments for and counter-arguments, on potential difficulties identified in SERVQUAL (Lee and Lin, 2005; Negash et al., 2003; Teas, 1993; Van Dyke et al., 1997, 1999). They further chronicle their misgivings on SERVQUAL’s conceptualization and empirical validations around: (1) operationalization of the perceived service quality as a difference or gap score, (2) ambiguity surrounding the expectations construct, and (3) unsuitability in the use of a single measure of service quality across different industries.

Counterarguments have been made in the literature elucidating the basis and conceptualization behind SERVQUAL and its five dimensions (Pitt et al., 1995; Klein et al., 2009). The current situation (i.e., arguments and counterarguments) in the literature on the applicability of the SERVQUAL instrument in business applications motivated our research. We argue for the inclusion of the tangibility dimension into the IS-adapted SERVQUAL framework because it is an important and relevant information service as well as operational dimension when conceptualized from an inherently abstract dimension like the other four dimensions. We believe that the physical features (tangibles) within IS operations extrinsically motivates customers to experience good service and, most importantly, value.

Indeed, IS-adapted SERVQUAL provides a means through which information service managers can differentiate themselves from their competition, gain a competitive advantage, and measure their effectiveness. Researchers assert that the ability to measure IS service systems, its operations and effectiveness in terms of its quality is a critical component (Gorla and Somers, 2014; Gorla, 2011; Kettinger and Smith, 2009; Seddon, Graeser, and Willcocks, 2002; Klein et al. 2009). In addition to determining whether the service delivery is meeting service expectations, IS-adapted SERVQUAL also provides a means to help managers improve functional areas within their service operations that are potentially damaging to their brand or less competitive than desired. As (Jiang et al. 2012) put it, the five dimensions of SERVQUAL (i.e., tangibles, reliability, responsiveness, assurance, and empathy) are the basis of much of the service quality research in IS. Further, Kettinger and Lee 1994 assert that the IS-adapted SERVQUAL instrument provides practical value to IS service managers in areas that need quality improvement. For information systems practitioners and researchers, the IS-adapted SERVQUAL instrument is important because it helps managers know how customers assess the quality of information service operations received based on the gap between what they expect and the perceived service delivery (Carr 2002; Ma, Pearson, and Tadisina, 2005; Kallweit et al., 2014). Moreover, it provides an avenue and platform to connect and engage with customers while soliciting feedback on their satisfaction and experiences during the service encounter. The literature provides support for the strength of the IS-adapted SERVQUAL instrument with its cross-cultural empirical examination and validation studies (Wu, Lin, and Cheng, 2009; Roses, Hoen, and Enrique, 2009; Negash et al., 2003; Devaraj, Fan, and Kohli, 2002; Jiang et al., 2003; Li, Tan, and Zie, 2002; Kettinger et al. 1995).

In the service quality area, difference score measures are typically conceptualized as the difference between a consumer’s expected or desired level of product performance measured prior to actual product usage and the consumer’s perception of actual performance after product usage (Parasuraman et al. 1985). Algebraically, the sum over all attributes of the differences is Σ*(Pi – Ei )*, where *Pi* is the perceived performance on attribute *i* and *Ei* is the expected or desired level of performance on that attribute. In the last two decades, research has seen studies developing models that utilize a match between two variables (Kettinger and Lee 1997; Klein et al. 2009; Petter et al. 2008; Pitt et al. 1995). For example, (Tesch et al. 2003) used a matching process for different stakeholders to study the perception and relationship of IS job performance, career satisfaction, and user satisfaction while (Bhattacherjee 2001) used a match of prior expectations and perceived performance of a system to study user satisfaction in the IS continuance model.

Many theoretical, conceptual, and empirical issues and reservations have been raised about the use of difference scores in testing models. In the last two decades, there were a considerable number of debates, arguments for and counter-arguments, on the potential difficulties identified in SERVQUAL (Klein et al. 2009; Van Dyke et al. 1997). Scholars have shown the difference scores to have low reliability (Johns 1981; Lord 1958). Other potential problems highlighted in the literature include: the inability of the difference scores to demonstrate the structure uniqueness from the perception component (i.e., discriminant validity) (Johns 1981; Wall and Payne 1973), variance restriction of the difference scores when one of the components used to create a difference score is consistently higher than the other (Wall and Payne 1973), and the instability in difference score-factor structure under varying contexts (Teas 1993).

Measurement errors resulting from diverse interpretation of expectations by respondents may be correlated across the levels of measurement (i.e. expected and perceived). In this case, the difference score will not be a reliable indicator of differences. This is because the observations are not independent when measurement errors are correlated. Commonly used parametric analyses assume observations to be independent and even other analyses that can work with non-independent data require that this independency be modeled. Therefore, there is a need to verify if error correlations exist and if they do, to devise a better method to understand the gap between the expected and perceived levels of service quality.

Thus, we question whether the validity of the IS-adapted SERVQUAL instrument has been comprehensively evaluated (i.e., for all traits/factors taking into account performance and expected service qualities, and potential measurement errors). Also, Yu et al (2008) contend that rating each variable on two service qualities may confound users to respond to items in each gap theory dimension in reference to the other. Hence, respondents may not express these differences clearly on an ordinal scale of measurement. Given that the two service qualities are correlated, (a) these correlations must be taken into account using a systematic approach while comprehensively validating the instrument; and (b) failure to include error correlations between items measured on the two service qualities will yield biased solutions (Byrne and Goffin 1993; Kenny and Kashy 1992). This debate and the questions arising from the literature give us the impetuses to ask:

1. Do respondents distinguish between the performance and expected service qualities of the same variable sufficiently well that measuring a gap is meaningful?
2. How are responses and their measurement errors on different service quality dimensions correlated and how can these correlations be systematically included in the model?

On a broader level, we are also interested in how the structure of the scores calculated based on gap theory in an instrument like IS-adapted SERVQUAL allows accounting for method effects. We show how this structure allows separating out the method variance that is attributed to many factors including the individual rater such as rater severity or personality traits (Geiser and Lockhart 2012), raters’ misunderstanding of the item, and raters’ understanding of the scale of measurement.

The first question about the ability of consumers to make distinctions is important because it directly addresses the validity of using gap scores to measure respondents’ evaluation of IS service quality. Computing gaps is valid only when the respondents distinguish between the two gap theory dimensions. The second question about the measurement error is equally important because the measurement errors of an item on the two gap theory dimensions may be correlated with each other. For instance, consider an individual with some measurement error on their rating of expectations of item (exp\_a1) “cloud-based service application platform provide prompt service delivery to users.” Perhaps this error is due to a slight misunderstanding of the question, rater severity, or an error in converting their true score to a 7-point Likert scale, or a combination of these. Logically, we would expect the same misunderstanding, severity, and/or conversion error to occur in their rating of item performance (perf\_a1) of the cloud-based service application platform’s prompt service delivery to its users. Therefore, correlation is expected among the errors associated with these items. Consider observed scores of items exp\_a1 and perf\_a1 that indicate the expected and actual performance levels of IS service quality, respectively. These scores are a sum of their true scores exp\_A1 and perf\_A1, respectively, and their error scores which are given as follows:

|  |  |  |
| --- | --- | --- |
|  | $$exp\\_a1=exp\\_A1+e\_{e1}$$ | (1) |
|  | $$perf\\_a1=perf\\_A1+e\_{p1}$$ | (2) |

The difference score exp\_a1 - perf\_a1 is a valid score if and only if the errors, $e\_{e1}$and $e\_{p1}$ are uncorrelated. This is because of the independence of observations assumption in most parametric analyses. Therefore, it is appropriate and necessary that these error correlations be included when evaluating validity. To our knowledge, the present study is the first of its kind to systematically include such error correlations and allow explaining the method variance in the IS-adapted SERVQUAL. As a result, we posit and test a more comprehensive measurement approach for use with gap scores that is valuable for use in future IS service quality systems and effectiveness research.

Our work proffers a better approach that is predicated on new methodological developments. Based on the framework of (Natesan and Aerts 2016), we use the multitrait multimethod (MTMM) framework to evaluate the validity of using the gap theory (i.e., expectations and performance) for IS-adapted SERVQUAL. MTMM models are used to simultaneously evaluate the construct validity of traits measured using different methods while also measuring method effects. Thus, the two dimensions formed two methods of the MTMM framework. We also evaluate the construct validity of using two gap theory dimensions to measure users’ rating of service quality of cloud-based service application platforms. Our study differs from that of Natesan and Aerts (2016) methodologically because the expected and performance levels of gap theory were administered at different points in time making these levels truly different methods of measurement.

**Multitrait Multimethod Models (MTMMs)**

CFA provides an elegant framework for examining the interrelationships between constructs and methods, models and adjusts for measurement errors, their correlations (uniquenesses), and error theories, in addition to examining construct validity - all within a single framework (Brown 2006). Unlike EFA, CFA is capable of modeling method effects (Brown 2006; Ketokivi and Schroeder 2004). Based on the seminal work of (Campbell and Fiske 1959), multitrait multimethod models (MTMMs) and their variants have been used to establish construct validity (e.g. (Widaman 2010)). MTMMs can be used to model different traits measured at different times or using different methods such as raters or modalities. Recently (Natesan and Aerts 2016) showed how MTMM allows validating the use of gap theory dimensions in instruments that are based on gap scores. MTMM can be used to evaluate construct validity by decomposing the method effects to evaluate both convergent and discriminant validity. In the present study there are two method factors, that is, the gap theory dimensions administered during different times, and five trait factors, that is, the information service quality dimensions. Each item indicates both a service quality dimension and the gap theory dimension it measures. The errors of the same base item that measures expected and performance levels of service quality were allowed to correlate.

The MTMM matrix is a symmetric correlation matrix of T × M rows; where T is the number of service quality dimensions and M is the number of gap theory dimensions administered during different times on which these service quality dimensions are measured. Modeling the error correlations between the expected and performance levels provides an elegant solution to partitioning out some method variance.

**METHODOLOGY**

Data for this study were collected using a self-administered questionnaire. Our survey instrument was the IS-adapted SERVQUAL instrument (see appendix A). All items were measured on a seven-point Likert type of scale (1 – strongly disagree to 7 – strongly agree). The population for the study was drawn from a major public university in the southeastern United States. An online survey method was used to collect data for the study. Our sample is relatively homogenous with regard to age because the subjects are enrolled in a required undergraduate course. However, within the age group range of our undergraduates we have variation in such demographics as major because the students come from across a large and diverse campus. As an incentive to participate in the survey, the instructors of these courses offered extra credit to the participants.

Distinct methods were used to collect different data at two time periods (i.e., administered same construct scale but measuring different gap theory dimensions using the same study participants at two different points in time) (Conway 1998). In the first stage of the survey, individuals answered questions relating to their expectations of a cloud-based service application platform prior to using it. The interval of time between scale administrations at two points in time is an important consideration in a study design such as ours. Researchers suggest that the length of time between scale administrations should be based on theoretical justifications in the literature and the stability of the construct(s) of interest (Le et al. 2010). If the constructs of interest are highly stable, then longer intervals between the first and second administration of the survey measures will not have any substantive change effect over that interval. However, shorter time interval periods are recommended for less stable construct(s) because substantive and essential changes in these constructs can occur over longer periods. Because our constructs are less stable and attitudinal in nature, the shorter time period was favored (Shaffer et al. 2016).

Therefore, a week after the expectation dimension of the survey was administered, the second gap theory dimension, that is, the perceived service quality was administered. In this second administration, participants were asked to rate their “perception” on the performance of selected cloud-based service application platform in the first scale administration. Following suggestions by Shaffer et al. (2016) to be mindful of environmental changes and its effects on studies data collection time periods, we made sure participants took both the first and the second surveys in the same computer lab to avoid any environmental changes. Examples of cloud-based service application platforms used for the study include Google Drive, Microsoft SkyDrive, Dropbox, and iCloud. The ability to collect meaningful and relevant data was another motivating reason for using these specific cloud-based service platforms. College students were deemed appropriate subjects for this study because they represent a significant part of the cloud application platform’s consumer population (Martensen 2007) and are in constant touch with peers through technology applications.

Out of the total 1,067 subjects who are enrolled in this core class, the first scale administration yielded 445 responses while the second survey administration produced 406 responses. Because data are collected on the same respondents over two time periods, we asked each participant during these two survey administrations to write down their emails. We therefore used these e-mails as an identifier in matching a respondent’s first survey with their second survey. In all, 328 e-mails (125 males and 203 females) were matched, indicating that these respondents completed both stages of the survey. The demographic information about the survey respondents is summarized in Table 1.

[**Insert Table 1 about here**]

The sample did not have univariate normality. Therefore, asymptotic covariance matrices (ACM) were used in model estimation (Jöreskog 1994). Because the scale of measurement is ordinal, we considered using polychoric correlations (PCM) instead of Pearson correlations (Kline 2015; Olsson 1979). However, complex models fitted using both ACM and PCM frequently run into admissibility issues. Moreover, the correlation matrices of continuous data and its corresponding categorical equivalent become closer when the number of categories is greater than 5 (Bollen and Barb 1981). We used a 7-point Likert scale. Therefore, we fitted the models using Pearson correlations instead of the PCM. Unweighted least squares method was used instead of maximum likelihood (Jöreskog 1990). When using the ACM, the admissibility condition may be set to zero and completely standardized solutions observed for mathematically inadmissible values. Model fit can be interpreted if and only if all completely standardized solution values lie within the interval of [-1, 1]. For all models, errors from the two methods (gap theory dimensions) for the same item were allowed to correlate using the appropriate syntax. In SIMPLIS (which was used for model fitting in this study) a sample syntax is “let the errors between Exp\_a1 and Perf\_a1 correlate”.

**MTMM model specification**

In the MTMM framework, we can decompose the covariance between traits as those due to method effects and “true” covariance. However, fully crossed MTMM models (all traits × all methods) evaluated using CFA often run into convergence problems and inadmissible solutions (Marsh 1989). Therefore, alternative models such as the correlated uniqueness (CU) model (Kenny 1976), the composite direct product model (Browne, 1984), and correlated traits correlated methods -1 (Eid 2000) have been proposed in literature.

In the following paragraphs, steps 1-3 list one possible logical sequence to instrument validation using MTMM: trait convergent validity, method convergent validity, and fully crossed trait and method factorial structure, respectively. Step 4 describes the MTMM and its alternative, the correlated uniqueness (CU) model. According to classical test theory, the score on a measured item $y$ is expressed as the sum of its true score $τ$ and error $ε$, as:

|  |  |  |
| --- | --- | --- |
|  | $$y= τ+ ε$$ |  (3) |

The factors that contribute to the score are the corresponding service quality factor score T and method factor score M. If $λ\_{T}$ and $λ\_{M}$ are the corresponding factor pattern coefficients, respectively, the relationship between the measured and latent scores is given as:

|  |  |  |
| --- | --- | --- |
|  | $$τ= λ\_{T}T+ λ\_{M}M$$ |  (4) |

For a sample of size $N$ with $I $items that are indicators of $T$ traits (service quality dimensions) and measured on $M$ methods (gap theory dimensions measured at different time-points), the models that can be used to address various validity issues are as follows. In the present study, there are 5 traits (tangibles, reliability, responsiveness, assurance, and empathy) and 2 methods (performance and expected service quality measured at different time-points). In the following paragraphs, steps 1-4 list a logical sequence to validating SERVQUAL using MTMM: service quality convergent validity, gap theory convergent validity, and fully crossed service quality and gap theory factorial structure, respectively.

**Step 1: Trait convergent validity.** Trait convergent validity for each method of measurement is supported when the set of $I$ items invokes the same conceptual framework in defining the latent construct in each method. Trait convergent validity can be tested by fitting a confirmatory factor model to all the items within each gap theory dimension.

**Step 2: Method convergent validity.** Method convergent validity is supported when the methods of measurement are clearly indicated by measuring the items on the different methods but traits are not taken into account. This provides evidence for using gap theory dimensions when administered during different times to measure gaps in service quality and indicates that the users were able to distinguish between the different gap theory dimensions. Validity of gap theory dimensions (methods) can be tested by fitting an *M* factor model to all the items without taking into consideration the latent traits.

**Step 3: Fully crossed convergent validity.** Correlations of fully crossed factors (*T* latent traits ×*M* methods) can be examined to evaluate convergent and discriminant validities, and method effects because the model decomposes each of these effects.

**Step 4: Multitrait multimethod models.** There are several models that can model multiple traits and methods. The most popular of them isthecorrelated traits correlated methods (CTCM) model. In the CTCM model the correlations between the trait factors and the methods are allowed to vary and the correlations across traits and methods are constrained to zero. In addition to being more parsimonious than the fully crossed model in step 3, this model also acknowledges the underlying structure that each item partitions out into measuring a gap theory dimension and a service quality dimension. Although a CTCM model may be fitted for the current type of instrumentation, the CTCM model requires at least three traits and three methods to be identified. Therefore, the CTCM was not fitted to the present data. Several types of constraints are placed on the CTCM model to give rise to other variants such as the perfectly correlated traits correlated methods model and the correlated traits uncorrelated methods model. Although the current data does not allow these models to be fitted, researchers may fit these models in gap theory-based instruments with at least three traits and three methods.

The correlated uniqueness (CU) model is a good and a widely recommended alternative to the CTCM model (Kenny and Kashy 1992; Marsh 1989; Marsh and Bailey 1991). In addition to being a parsimonious alternative and working well for instruments with less than three traits or three methods, the CU model does not run into convergence issues like the CTCM model. In the CU model, the measurement errors of the same items across different methods are allowed to correlate. Errors cannot be correlated across items measured by the same method (Kenny, 2004, p.190).

Models in steps 1-4 were fitted to the data. Because we had only two methods of measurement, the CU model was fit to the data in the present study. The CFA and part of the correlated uniqueness models are shown in Figures 1 and 2, respectively. Factor pattern and structure coefficients were computed for the model of best fit. Structure coefficient is the correlation between a latent variable and an observed variable. This is sometimes referred to as indicator-factor correlation. In CFA, reporting structure coefficients is important because low value of structure coefficient of the item that does not load on a factor indicates support for discriminant validity (Courville and Thompson 2001).

**[Insert Figures 1 and 2 about here]**

Due to the non-normality of the data, Satorra-Bentler scaled χ² ($χ\_{SB}^{2}$) statistics are interpreted. The width of the RMSEA confidence interval was also examined to see if the sample size was adequate to obtain precise results. Model modifications were made based on standardized residuals and theory, and fit indices were checked. Modifications were considered for paths that had absolute values of standardized residuals greater than 1.96 which corresponds to a p-value of 0.05. However, paths were not added if the theory did not support such model modification. For instance, errors of two items that indicated different service quality dimensions and were measured on different gap theory dimensions were not correlated. The stem-leaf display of standardized residuals was observed for symmetry in each model. An excess of residuals on either side of the plot (positive or negative) may indicate systematic underestimation or overestimation, respectively (Jöreskog 1994). The simplest model was retained if it fit the theory and the rest of the fit indices indicated good fit. Model fit was also determined based on CFI (> 0.95, (Hu and Bentler 1999)), RMSEA, SRMR (< 0.08, (Hu and Bentler 1999)), and PSRMR (Corten et al. 2002). RMSEA < 0.05 indicated good fit and 0.05 < RMSEA < 0.08 indicated medium but acceptable fit (Browne and Cudeck 1993). The parsimony standardized root mean squared residual (PSRMR) adjusts the SRMR for loss of degrees of freedom and presents the average residual per degree of freedom.

**RESULTS**

For all the retained models the exact-fit test was not statistically significant as seen in Table 2. Standardized residuals for the CFA models within each method and the CU model were less than 1.96 which corresponds to a p-value of 0.05. In all models, error correlations were shown to be possible sources of model misfit because they had the highest standardized residuals. The CFA models within each method, the CU model, and the method convergent validity model passed the model fit criteria for good model fit and therefore, were retained. Of the models that accounted for both trait and method variance, the CU model had the lowest$χ\_{SB}^{2}$**,** SRMR, and PSRMR. The largest standardized residual for this model was 0.74.

**[Insert Table 2]**

The CFA models within each method fit the data well. This indicates support for the 5-factor structure. Although some factors were more highly correlated than others, this model was retained because this was the only model that fit both the methods well. A single factor model also fit the data well but had high standardized residuals (up to 7.38). These residuals indicated correlation between items that indicated the same factor in the 5-factor model. The 10-factor model with fully crossed traits and methods could not be fit because of non-positive-definite covariance matrix. When the admissibility was set to off, some of the standardized solutions were greater than 1. Therefore, this model was rejected. As discussed before, this is a common problem when fitting a complex model to non-normal data (Jöreskog and Sörbom 1996).

Trait convergent validity model was not retained because it failed the SRMR and GFI cutoff criteria. There were several error correlations suggested by LISREL to improve model fit. All of these involved correlating the errors of the same item measured across different methods. Adding these error correlations would have converted the model into a CU model. Therefore, these model modifications were not done. However, the important point to note here is that the data suggested that *method correlations were necessary to obtain better model fit*.

The method convergent validity model fits the data well indicating support for participants discerning the difference between the methods of measurement in gap theory measured at different time-points. Like the fully crossed model, the CU model also yielded non-positive-definite matrix initially. After the admissibility was set to off, it yielded good model fit with all standardized solutions less than 1. Therefore, this model was retained.

The CU model fits the data better than the trait and method convergent validity models. This indicates support for the five-factor structure of SERVQUAL while showing that the participants discerned the gap theory dimensions. The error correlations between items belonging to the same method ranged from 0.19 to 0.79 for the expected dimension and from 0.10 to 0.92 for performance dimension. This indicates support for the use of gap theory dimensions because items within the same gap theory dimension are correlated with each other. In this way, we are able to partition out the method variance.

**[Insert Table 3]**

The factor correlations for the five service quality dimensions in the CU model: tangibles, reliability, responsiveness, assurance, and empathy ranged from 0.0.28 to 0.66 (Table 3). The structure coefficients were sometimes the same as or very close to the factor coefficients because of high factor correlations. This is far from an ideal situation because there is a threat to discriminant validity in such cases. For example, tangibles factor was perfectly and almost perfectly correlated with reliability and responsiveness, respectively. However, no other factor structure fit well within both methods. Furthermore, in many past studies the number of factors was collapsed (Van Dyke et al. 1997; Van Dyke et al. 1999) with a different number of factors being suggested in different applications. The results of this study suggest that the high correlations among factors are a likely source of the rationale for collapsing the number of factors. In this research study we retained five factors because that is consistent with the SERVQUAL base theory and retaining five factors shows a more general application of the methodology than would be apparent with a specialized factor structure such as suggested in prior studies with a specific application.

***Final synopsis:*** We found the IS-adapted SERVQUAL instrument not perfect, especially with respect to the underlying 5-factor structure. This is the reason for several undesirably high trait factor correlations that indicate threats to discriminant validity. However, this analysis indicates that the MTMM framework can effectively decompose the method effects to investigate whether the respondents can distinguish between the gap theory dimensions. Secondly, error correlations between measured variables are important to take into account because of the underlying structure of gap scores. Some very high error correlations, as seen in the theta-delta matrix, indicate that these scores cannot be subtracted from each other to form a valid difference score. The trait convergent validity model did not include error correlations between the same items measured using different methods had inadequate fit. This further demonstrates the need for a comprehensive framework such as the one presented here to evaluate the validity of instruments that use gap scores. As such this work has implications for numerous instruments that use gap theory, including the use of SERVQUAL in service operations.

**[Insert Table 4]**

**DISCUSSION**

This study relies upon and extends the work of (Natesan and Aerts 2016) and shows that, for IS service quality system and effectiveness research, the MTMM framework can effectively decompose the method effects to investigate whether the respondents can distinguish between the gap theory dimensions in the SERVQUAL instrument. In many studies, the match is a computed score measure calculated using the difference between the two variables under study. Using difference scores in developing models has produced conflicting and inconsistent results. In addition, these difference scores or indirect measures have led to erroneous conclusions that are largely due to the inherent false implicit assumptions about its philosophy and data characteristics.

While some researchers identify dimensionality issues when using difference score measure, others also express issues concerning reliability and validity of these indirect measures. In response to such anomalies and concerns associated with the difference score measure, some scholars have moved on to develop models with the SERVQUAL instrument using a more direct measure approach. Here, they eliminate one of the dimensions (i.e., the expectation dimension) and use only the performance dimension for their work. Though consistent results are produced with studies using this approach, questions have been raised regarding the theoretical underpinnings of such an approach (Klein et al. 2009). Moreover, this approach is potentially cumbersome when more than two gap theory dimensions are involved. In all, both sides of the research divide bring value to the intellectual discourse and our work helps bridge both sides. The current study contributes to the IS service quality literature because it provides an opportunity to use gap scores in a manner consistent with the theoretical foundations of the models, such as the use of SERVQUAL to measure IS service system quality and effectiveness, as well as resolving some of the associated methodological limitations. In the context of IS-adapted SERVQUAL we offer a systematic procedure researchers can use to comprehensively evaluate the validity of instruments that use a match between two variables with the inclusion of error correlations. Largely from a measurement model perspective, our work delineates how well respondents are able to distinguish between the gap theory dimensions of the SERVQUAL instrument. Most importantly we show that gap scores obtained simply by subtracting item scores obtained from different gap theory dimensions are not valid because the two scores are not independent of each other.

From a dimensionality standpoint, research suggests the unstable structure of the 5-factor structure of SERVQUAL across industries (Van Dyke et al. 1997; Van Dyke et al. 1999). This is given much prominence in recent studies where the IS-adapted SERVQUAL instrument collapses into a four-factor model (Lee and Lin 2005; Negash et al. 2003; Van Dyke et al. 1999). The collapse of the tangibility dimension is attributed to the fact that this dimension is either difficult to conceptualize or operationalize in the IS domain. Although the 5-factor structure had discriminant validity issues, the 4-factor structure did not even possess adequate model fit for retention.

Our study’s objectives were to assess if respondents can distinguish between the two gap theory dimensions (i.e., expectation and performance) and to find out how the correlation between the measurement errors on these dimensions can be systematically included in the model to partition method effects. In our analysis, we used the MTMM framework, the CU model, in particular, to account for method effects. There is significant support for method effects as shown by the suggested model paths in the trait convergent validity model. Additionally, the good model fit for the CU model and the medium to high error correlations between items within the same gap theory dimension show that respondents are able to distinguish between the two gap theory dimensions of the SERVQUAL measurement. Although medium-large factor coefficients indicate convergent validity, the extremely large structure coefficients indicate a threat to discriminant validity.

One logic of using two different gap theory dimensions is to provide the user with a scale of reference when rating the same construct on different methods. The scores thus obtained must then reflect this assumption. This research has the potential to change the approach to using SERVQUAL and as such makes an important contribution to theory that is likely to influence practice. It is our hope to see more IS, retail and consumer service researchers adopt the measurement approach we posit and test because (a) it refines the definition of the measures by incorporating error correlations and therefore, (b) different results will be obtained when applying our approach within a different context. Most importantly, our approach shows that the use of difference scores as a gap indicator can be highly unreliable. This is because the model could not be fitted without accounting for error correlations between the same base items from different methods. As future work develops, the results of this research offer an opportunity to advance the measurement of service quality and the associated practical decisions across a variety of business disciplines.

**LIMITATIONS AND FUTURE RESEARCH**

Notwithstanding our contributions, we note several limitations of our study. First, in our illustration, we have used the adapted IS-adapted SERVQUAL instrument as the basis to assess respondents’ ability to distinguish between the two gap theory dimensions (i.e., expectation and performance). We note that these respondents are students and that they may not be able to fully engage in their experience with the use of cloud based service platforms for task related performance and as such working professionals should be targeted to assess the distinction between expectation and performance. Second, our study applies to situations where respondents are asked about their experiences with cloud-based service platforms from multiple services providers. There is a lack of parallelism among the individual’s orientation of IS SERVQUAL since a subject (respondent) could be responding with respect to Google’s service and another Microsoft’s service. Future research in such situations that canvas a broad range of diverse services from multiple service providers, will need to examine specific services and differences in service quality experiences of such information platforms.

Third, we could not find a common underlying factor structure that showed adequate indications of convergent and discriminant validities for the individual CFA models within each gap theory dimension and the CU model. Finally, it will be interesting to compare the service quality measure of managers with those of customers by collecting data from these two groups to ascertain a solution to the consensus of service quality measurement. Although we cannot predict the invariance of such an approach across different groups, the current analysis can be extended to test invariance of the latent variable across the different groups before making group comparisons.

REFERENCES

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly 25*(3), 351-370.

Boakye, K. G., McGinnis, T., Prybutok, V. R., & Paswan, A. K. (2014). Development of a service continance model with IT service antecedents. *Journal of Retailing and Consumer Services*, *21*(5), 717-724.

Bollen, K. A., & Barb, K. H. (1981). Pearson's R and coarsely categorized measures, *American Sociological Review*, *46*(2), 232-239.

Brown, T. (2006). CFA with equality constraints, multiple groups, and mean structures. In T. Brown (Ed.), *Confirmatory factor analysis for applied research*, . 236-319. New York: Guildford.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136-162), Newbury Park, CA: Sage.

Byrne, B. M., & Goffin, R. D. (1993). Modeling MTMM data from additive and multiplicative covariance structures: An audit of construct validity concordance, *Multivariate Behavioral Research* *28*(1), 67-96.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix, *Psychological Bulletin,* *56*(2), 81-105.

Carr, C. L. (2002). A psychometric evaluation of the expectations, perceptions, and difference‐scores generated by the is‐adapted servqual instrument, *Decision Sciences*, 33(2), 281-296.

Chase, R. B., & Apte, U. M. (2007). A history of research in service operations: What's the big idea?, *Journal of Operations Management*, 25 (2), 375-386

Conway, J. M. (1998). Estimation and uses of the proportion of method variance for multitrait-multimethod data, *Organizational Research Methods* (1:2), 209-222.

Corten, I. W., Saris, W. E., Coenders, G., van der Veld, W., Aalberts, C. E., & Kornelis, C. (2002). Fit of different models for multitrait-multimethod experiments, *Structural Equation Modeling* 9(2), 213-232.

Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement,* *61*(2), 229-248.

Devaraj, S., Fan, M., & Kohli, R. (2002). Antecedents of b2c channel satisfaction and preference: Validating e-commerce metrics, *Information System Research,* 13(3), 316-333.

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions, *Psychometrika,* 65(2), 241-261.

Geiser, C., & Lockhart, G. (2012). A comparison of four aroaches to account for method effects in latent state–trait analyses. *Psychological Methods,* *17*(2), 255-283.

Gorla, N., & Somers, T. M. (2014). The impact of IT outsourcing on information systems success. *Information & Management,* *51*(3), 320-335.

Gorla, N. (2011). An assessment of information systems service quality using SERVQUAL+. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *42*(3), 46-70.

Grover, V., Cheon, M. J., & Teng, J. T. (1996). The effect of service quality and partnership on the outsourcing of information systems functions. *Journal of Management Information Systems,* *12*(4), 89-116.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal,* *6*(1), 1-55.

Jiang, J. J., Klein, G., & Crampton, S. M. (2000). A note on servqual reliability and validity in information system service quality measurement, *Decision Sciences,* *31*(3), 725-744.

Jiang, J. J., Klein, G., Parolia, N., & Li, Y. (2012). An analysis of three servqual variations in measuring information system service quality. *Electronic Journal Information Systems Evaluation,* *15*(2), 149-162.

Jiang, J.J., Klein, G., Roan, J. & Lin, J.T. (2001). IS service performance: Self-perceptions and user perceptions. *Information & Management*, *38*(8), 499-506.

Jiang, J. J., Klein, G., Tesch, D., & Chen, H. G. (2003). Closing the user and provider service: quality gap, *Communcation of the ACM,* 46 (2), 72-76.

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance,* *27*(3), 443-463.

Jöreskog, K. G. (1990(. New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity,* *24*(4), 387-404.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix, *Psychometrika,* *59*(3), 381-389.

Jöreskog, K. G., & Sörbom, D. (1996(. *Prelis 2 user's reference guide: A program for multivariate data screening and data summarization: A preprocessor for lisrel*. Scientific Software International.

Kallweit, K., Spreer, P., & Toporowski, W. (2014). Why do customers use self-service information technologies in retail? The mediating effect of perceived service quality. *Journal of Retailing and Consumer Services*, *21*(3), 268-276.

Kenny, D. A. (1976). An empirical alication of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology,* *12*(3), 247-252.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin,* *112*(1), 165-172.

Ketokivi, M. A., & Schroeder, R. G. 2004. Strategic, structural contingency and institutional explanations in the adoption of innovative manufacturing practices. *Journal of Operations Management,* *22*(1), 63-89.

Kettinger, W.J. & Smith, J. (2009). Understanding the consequences of information systems service quality on IS service reuse. *Information & Management*, *46*(6), 335-341.

Kettinger, W. J., & Lee, C. C. (1994). Perceived service quality and user satisfaction with the information services function. *Decision Sciences,* 25(5‐6), 737-766.

Kettinger, W. J., & Lee, C. C. (1997). Pragmatic perspectives on the measurement of information systems service quality. *MIS Quarterly, 21*(2), 223-240.

Kettinger, W. J., Lee, C. C., & Lee, S. (1995). Global measures of information service quality: A cross‐national study. *Decision Sciences,* *26*(5), 569-588.

Klein, G., Jiang, J. J., & Cheney, P. (2009). Resolving difference score issues in information systems research. *MIS Quarterly*, *33*(4), 811-826.

Kline, R. B. 2015. *Principles and Practice of Structural Equation Modeling* (4th ed.). New York: Guilford Press.

Le, H., Schmidt, F. L., Harter, J. K., & Lauver, K. J. (2010). The problem of empirical redundancy of constructs in organizational research: An empirical investigation. *Organizational Behavior and Human Decision Processes, 112*(2), 112-125.

Lee, G.-G., & Lin, H.-F. (2005). Customer perceptions of e-service quality in online shoing. *International Journal of Retail & Distribution Management,* *33*(2), 161-176.

Li, Y. N., Tan, K. C., & Xie, M. (2002). Measuring web-based service quality. *Total Quality Management*, 13(5), 685-700.

Lord, F. M. 1958. The utilization of unreliable difference scores. *Journal of Educational Psychology,* 19(3), 150-152.

Ma, Q., Pearson, J.M. & Tadisina, S., (2005). An exploratory study into factors of service quality for alication service providers. *Information & Management,* *42*(8), 1067-1080.

Malhotra, M. K., & Sharma, S. (2008). Measurement equivalence using generalizability theory: An examination of manufacturing flexibility dimensions. *Decision Sciences,* *39*(4), 643-669.

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement,* *13*(4), 335-361.

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement,* 15(1), 47-70.

Martensen, A. (2007). Tweens' satisfaction and brand loyalty in the mobile phone market. *Young Consumers,* *8*(2), 108-116.

Natesan, P., & Aerts, X. (2016). Can library users distinguish between minimum, perceived, and desired levels of service quality? Validating Libqual+® using multitrait multimethod analysis. *Library & Information Science Research,* 38(1), 30-38.

Negash, S., Ryan, T., & Igbaria, M. (2003). Quality and effectiveness in web-based customer suort systems. *Information & Management,* *40*(8), 757-768.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika,* *44*(4), 443-460.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *The Journal of Marketing*, *49*(4), 41-50.

Petter, S., DeLone, W., & McLean, E. (2008). Measuring information systems success: Models, dimensions, measures, and interrelationships. *European Journal of Information Systems,* *17*(3), 236-263.

Pitt, L. F., Watson, R. T., & Kavan, C. B. (1995). Service quality: A measure of information systems effectiveness. *MIS Quarterly, 19*(2), 173-187.

Rosene, F. (2003). Complacency and service quality: An overlooked condition in the GAP model. *Journal of Retailing and Consumer Services*, *10*(1), 51-55.

Roses, L. K., Hoppen, N., & Henrique, J. L. (2009). Management of perceptions of information technology service quality, *Journal of Business Research*, 62 (9), 876-882.

Seddon, P. B., Graeser, V., & Willcocks, L. P. (2002). Measuring organizational IS effectiveness: an overview and update of senior management perspectives. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *33*(2), 11-28.

Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods,* *19*(1), 80-110.

Teas, R. K. (1993). Expectations, performance evaluation, and consumers' perceptions of quality, *The Journal of Marketing*, *57*(4), 18-34.

Tesch, D., Jiang, J. J., & Klein, G. (2003). The impact of information system personnel skill discrepancies on stakeholder satisfaction. *Decision Sciences,* *34*(1), 107-129.

Van Dyke, T. P., Kappelman, L. A., & Prybutok, V. R. (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. A. 1999. Cautions on the use of the SERVQUAL measure to assess the quality of information systems services. *Decision Sciences,* 30(3), 877-891.

Wall, T. D., & Payne, R. (1973). Are deficiency scores deficient? *Journal of Applied Psychology,* 58(3), 322-326.

Widaman, K. F. (2010). Multitrait-multimethod analysis. In G.R. Hancock, & R.O. Mueller (Eds.). *The reviewer's guide to quantitative methods in the social sciences,* Oxford, UK: Routledge, 299-314.

Wu, W., Lin, B., & Cheng, C. F. (2009). Evaluating online auction strategy: a theoretical model and empirical exploration, *Journal of Computer Information Systems*, 49 (3), 22-30.

Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1990). Delivering quality service: balancing customer perceptions and expectations. Free Press, New York, NY.

**Appendix A. Measurement items.**

**DIRECTIONS:** We want your opinion and feeling on your quality expectations on using a cloud-based service application. For each of the following statements, please indicate your agreement or disagreement with your level of expectations on what a cloud-based service application platform should deliver. There are no right or wrong answers. Please respond to each of the following statements using a scale from 1 (strongly disagree) to 7 (strongly agree). Select 4 if you feel that you are neutral or indifferent.

I expect a quality cloud-based service application platform to **\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1. Have an up-to-date hardware and software. (T)
 |  |  |  |  |  |  |  |
| 1. Be visually appealing. (T)
 |  |  |  |  |  |  |  |
| 1. Be neat and professionally designed. (T)
 |  |  |  |  |  |  |  |
| 1. Reflect the kind of service provided. (T)
 |  |  |  |  |  |  |  |
| 1. Provide services as promised. (R)
 |  |  |  |  |  |  |  |
| 1. Provide avenues in solving user’s problems. (R)
 |  |  |  |  |  |  |  |
| 1. Be dependable. (R)
 |  |  |  |  |  |  |  |
| 1. Perform service right the first time. (R)
 |  |  |  |  |  |  |  |
| 1. Be error-free. (R)
 |  |  |  |  |  |  |  |
| 1. Be always available to users. (Rs)
 |  |  |  |  |  |  |  |
| 1. Provide prompt delivery to users. (Rs)
 |  |  |  |  |  |  |  |
| 1. Be willing to help users. (Rs)
 |  |  |  |  |  |  |  |
| 1. Show readiness to respond to user’s requests. (Rs)
 |  |  |  |  |  |  |  |
| 1. Instill confidence in users. (A)
 |  |  |  |  |  |  |  |
| 1. Provide secured transaction platform for users. (A)
 |  |  |  |  |  |  |  |
| 1. Be considerate with users. (A)
 |  |  |  |  |  |  |  |
| 1. Provide information needed to perform a task. (A)
 |  |  |  |  |  |  |  |
| 1. Provide distinct attention to users. (E)
 |  |  |  |  |  |  |  |
| 1. Have convenient hours of service operations. (E)
 |  |  |  |  |  |  |  |
| 1. Provide personalized attention to users. (E)
 |  |  |  |  |  |  |  |
| 1. Serve users’ interest well. (E)
 |  |  |  |  |  |  |  |
| 1. Understand specific needs of users. (E)
 |  |  |  |  |  |  |  |

Note: T = Tangibles, R = Reliability, Rs = Responsiveness, A = Assurance, E = Empathy

**DIRECTIONS:** We want your opinion and experience on the performance of your use of a cloud-based service application relative to your prior expectations**.** For each of the following statements, please indicate your agreement or disagreement with your performance level of your experience in using cloud-based application service platforms. There are no right or wrong answers. Please respond to each of the following statements using a scale from 1 (strongly disagree) to 7 (strongly agree). Select 4 if you feel that you are neutral or indifferent.

How strongly do you agree or disagree with the following statements regarding the performance level of your preferred cloud-based service application platform relative to your initial expectations?

My preferred cloud-based service application platform \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1. Has an up-to-date hardware and software. (T)
 |  |  |  |  |  |  |  |
| 1. Is visually aealing. (T)
 |  |  |  |  |  |  |  |
| 1. Is neat and professionally designed. (T)
 |  |  |  |  |  |  |  |
| 1. Reflects the kind of service provided. (T)
 |  |  |  |  |  |  |  |
| 1. Provides services as promised. (R)
 |  |  |  |  |  |  |  |
| 1. Provides avenues in solving user’s problems. (R)
 |  |  |  |  |  |  |  |
| 1. Is dependable. (R)
 |  |  |  |  |  |  |  |
| 1. Performs service right the first time. (R)
 |  |  |  |  |  |  |  |
| 1. Is error-free. (R)
 |  |  |  |  |  |  |  |
| 1. Is always available to users. (Rs)
 |  |  |  |  |  |  |  |
| 1. Provides prompt delivery to users. (Rs)
 |  |  |  |  |  |  |  |
| 1. Is willing to help users. (Rs)
 |  |  |  |  |  |  |  |
| 1. Shows readiness to respond to user’s requests. (Rs)
 |  |  |  |  |  |  |  |
| 1. Instills confidence in users. (A)
 |  |  |  |  |  |  |  |
| 1. Provides secured transaction platform for users. (A)
 |  |  |  |  |  |  |  |
| 1. Is considerate with users. (A)
 |  |  |  |  |  |  |  |
| 1. Provides information needed to perform a task. (A)
 |  |  |  |  |  |  |  |
| 1. Provides distinct attention to users. (E)
 |  |  |  |  |  |  |  |
| 1. Has convenient hours of service operations. (E)
 |  |  |  |  |  |  |  |
| 1. Provides personalized attention to users. (E)
 |  |  |  |  |  |  |  |
| 1. Serves users’ interest well. (E)
 |  |  |  |  |  |  |  |
| 1. Understands specific needs of users. (E)
 |  |  |  |  |  |  |  |

Note: T = Tangibles, R = Reliability, Rs = Responsiveness, A = Assurance, E = Empathy

|  |
| --- |
| **Appendix B.** Theta-delta matrix for expected level of service quality |
|  | T1 | T2 | T3 | T4 | R1 | R2 | R3 | R4 | R5 | Rs1 | Rs2 | Rs3 | Rs4 | A1 | A2 | A3 | A4 | E1 | E2 | E3 | E4 | E5 |
| T1 | 0.72 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| T2 | 0.32 | 0.76 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| T3 | 0.45 | 0.37 | 0.74 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| T4 | 0.43 | 0.38 | 0.48 | 0.68 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Rl1 | 0.44 | 0.26 | 0.51 | 0.45 | 0.72 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Rl2 | 0.35 | 0.25 | 0.39 | 0.36 | 0.37 | 0.71 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Rl3 | 0.45 | 0.30 | 0.50 | 0.39 | 0.57 | 0.40 | 0.76 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Rl4 | 0.35 | 0.26 | 0.42 | 0.33 | 0.46 | 0.26 | 0.51 | 0.74 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Rl5 | 0.29 | 0.20 | 0.32 | 0.26 | 0.30 | 0.30 | 0.35 | 0.40 | 0.74 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Rs1 | 0.45 | 0.26 | 0.46 | 0.41 | 0.46 | 0.39 | 0.50 | 0.46 | 0.38 | 0.77 |  |  |  |  |  |  |  |  |  |  |  |  |
| Rs2 | 0.41 | 0.26 | 0.42 | 0.34 | 0.45 | 0.34 | 0.48 | 0.45 | 0.33 | 0.53 | 0.67 |  |  |  |  |  |  |  |  |  |  |  |
| Rs3 | 0.36 | 0.2 | 0.33 | 0.28 | 0.37 | 0.38 | 0.40 | 0.25 | 0.26 | 0.36 | 0.35 | 0.64 |  |  |  |  |  |  |  |  |  |  |
| Rs4 | 0.39 | 0.28 | 0.39 | 0.32 | 0.42 | 0.42 | 0.45 | 0.33 | 0.35 | 0.40 | 0.40 | 0.53 | 0.70 |  |  |  |  |  |  |  |  |  |
| A1 | 0.19 | 0.31 | 0.27 | 0.30 | 0.24 | 0.29 | 0.32 | 0.21 | 0.29 | 0.24 | 0.23 | 0.28 | 0.37 | 0.69 |  |  |  |  |  |  |  |  |
| A2 | 0.42 | 0.31 | 0.40 | 0.40 | 0.41 | 0.35 | 0.41 | 0.41 | 0.35 | 0.43 | 0.39 | 0.41 | 0.46 | 0.33 | 0.79 |  |  |  |  |  |  |  |
| A3 | 0.34 | 0.31 | 0.40 | 0.35 | 0.41 | 0.44 | 0.44 | 0.38 | 0.32 | 0.44 | 0.43 | 0.47 | 0.47 | 0.34 | 0.45 | 0.71 |  |  |  |  |  |  |
| A4 | 0.44 | 0.34 | 0.48 | 0.41 | 0.50 | 0.46 | 0.54 | 0.45 | 0.39 | 0.51 | 0.49 | 0.48 | 0.55 | 0.34 | 0.38 | 0.48 | 0.77 |  |  |  |  |  |
| E1 | 0.28 | 0.28 | 0.27 | 0.35 | 0.29 | 0.33 | 0.29 | 0.25 | 0.29 | 0.30 | 0.29 | 0.33 | 0.37 | 0.38 | 0.39 | 0.41 | 0.45 | 0.64 |  |  |  |  |
| E2 | 0.35 | 0.29 | 0.38 | 0.35 | 0.41 | 0.35 | 0.39 | 0.31 | 0.30 | 0.37 | 0.34 | 0.39 | 0.48 | 0.27 | 0.47 | 0.40 | 0.40 | 0.31 | 0.70 |  |  |  |
| E3 | 0.26 | 0.29 | 0.27 | 0.29 | 0.21 | 0.30 | 0.25 | 0.20 | 0.27 | 0.24 | 0.23 | 0.31 | 0.31 | 0.37 | 0.33 | 0.31 | 0.34 | 0.33 | 0.26 | 0.63 |  |  |
| E4 | 0.37 | 0.33 | 0.38 | 0.40 | 0.33 | 0.35 | 0.38 | 0.29 | 0.33 | 0.37 | 0.35 | 0.38 | 0.39 | 0.43 | 0.47 | 0.37 | 0.43 | 0.35 | 0.34 | 0.41 | 0.70 |  |
| E5 | 0.26 | 0.24 | 0.26 | 0.26 | 0.21 | 0.29 | 0.29 | 0.22 | 0.27 | 0.26 | 0.26 | 0.33 | 0.33 | 0.39 | 0.33 | 0.35 | 0.32 | 0.25 | 0.25 | 0.33 | 0.44 | 0.56 |

T = Tangibles, R = Reliability, Rs = Responsiveness, A = Assurance, E = Empathy

|  |
| --- |
| **Appendix B.** Theta-delta matrix for performance level of service quality |
|  | T1 | T2 | T3 | T4 | R1 | R2 | R3 | R4 | R5 | Rs1 | Rs2 | Rs3 | Rs4 | A1 | A2 | A3 | A4 | E1 | E2 | E3 | E4 | E5 |
| T1 | 0.85 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| T2 | 0.30 | 0.87 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| T3 | 0.42 | 0.47 | 0.89 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| T4 | 0.33 | 0.29 | 0.43 | 0.81 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Rl1 | 0.32 | 0.28 | 0.41 | 0.49 | 0.90 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Rl2 | 0.25 | 0.25 | 0.37 | 0.32 | 0.40 | 0.83 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Rl3 | 0.29 | 0.23 | 0.36 | 0.34 | 0.42 | 0.41 | 0.90 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Rl4 | 0.32 | 0.25 | 0.38 | 0.36 | 0.45 | 0.48 | 0.52 | 0.92 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Rl5 | 0.15 | 0.17 | 0.21 | 0.19 | 0.26 | 0.32 | 0.28 | 0.37 | 0.82 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Rs1 | 0.32 | 0.30 | 0.41 | 0.33 | 0.41 | 0.36 | 0.48 | 0.55 | 0.30 | 0.90 |  |  |  |  |  |  |  |  |  |  |  |  |
| Rs2 | 0.28 | 0.28 | 0.32 | 0.29 | 0.37 | 0.27 | 0.36 | 0.43 | 0.32 | 0.55 | 0.89 |  |  |  |  |  |  |  |  |  |  |  |
| Rs3 | 0.24 | 0.23 | 0.27 | 0.24 | 0.28 | 0.31 | 0.30 | 0.36 | 0.10 | 0.32 | 0.30 | 0.68 |  |  |  |  |  |  |  |  |  |  |
| Rs4 | 0.24 | 0.32 | 0.31 | 0.32 | 0.35 | 0.32 | 0.36 | 0.45 | 0.17 | 0.35 | 0.32 | 0.45 | 0.83 |  |  |  |  |  |  |  |  |  |
| A1 | 0.20 | 0.29 | 0.33 | 0.25 | 0.24 | 0.27 | 0.29 | 0.32 | 0.18 | 0.28 | 0.25 | 0.30 | 0.41 | 0.82 |  |  |  |  |  |  |  |  |
| A2 | 0.30 | 0.31 | 0.41 | 0.31 | 0.37 | 0.36 | 0.33 | 0.43 | 0.27 | 0.37 | 0.37 | 0.40 | 0.40 | 0.42 | 0.91 |  |  |  |  |  |  |  |
| A3 | 0.18 | 0.27 | 0.36 | 0.28 | 0.37 | 0.28 | 0.36 | 0.46 | 0.16 | 0.36 | 0.31 | 0.35 | 0.47 | 0.45 | 0.46 | 0.85 |  |  |  |  |  |  |
| A4 | 0.22 | 0.16 | 0.3 | 0.29 | 0.36 | 0.26 | 0.35 | 0.35 | 0.17 | 0.30 | 0.33 | 0.30 | 0.39 | 0.30 | 0.42 | 0.42 | 0.85 |  |  |  |  |  |
| E1 | 0.19 | 0.24 | 0.23 | 0.24 | 0.32 | 0.22 | 0.29 | 0.38 | 0.21 | 0.32 | 0.28 | 0.33 | 0.35 | 0.31 | 0.37 | 0.44 | 0.33 | 0.81 |  |  |  |  |
| E2 | 0.25 | 0.18 | 0.28 | 0.22 | 0.33 | 0.29 | 0.27 | 0.42 | 0.18 | 0.36 | 0.38 | 0.29 | 0.33 | 0.26 | 0.39 | 0.39 | 0.31 | 0.31 | 0.84 |  |  |  |
| E3 | 0.25 | 0.23 | 0.24 | 0.18 | 0.27 | 0.35 | 0.30 | 0.38 | 0.21 | 0.30 | 0.23 | 0.35 | 0.28 | 0.42 | 0.32 | 0.40 | 0.36 | 0.45 | 0.34 | 0.86 |  |  |
| E4 | 0.25 | 0.28 | 0.23 | 0.30 | 0.33 | 0.28 | 0.31 | 0.39 | 0.16 | 0.33 | 0.29 | 0.28 | 0.37 | 0.33 | 0.39 | 0.41 | 0.35 | 0.35 | 0.36 | 0.40 | 0.86 |  |
| E5 | 0.22 | 0.20 | 0.26 | 0.33 | 0.33 | 0.30 | 0.34 | 0.31 | 0.18 | 0.26 | 0.27 | 0.31 | 0.32 | 0.34 | 0.35 | 0.43 | 0.35 | 0.30 | 0.27 | 0.37 | 0.46 | 0.79 |

T = Tangibles, R = Reliability, Rs = Responsiveness, A = Assurance, E = Empathy

|  |  |
| --- | --- |
| **Table 1** |   |
| Demographic distribution of survey respondents.  |   |
| Gender | Percent of sample |
| Male | 38.1% |
| Female | 61.9% |
|  |  |
| Age |  |
| 18-20 | 54.4% |
| 21-25 | 40.2% |
| 26-30 | 3.0% |
| >30 | 2.4% |
| Cloud service platforms |  |
| Dropbox | 9.1% |
| Google Drive | 59.3% |
| iCloud | 27.0% |
| Microsoft OneDrive | 4.3% |
| Others | 0.3% |
|  |  |
| Monthly usage |  |
| Never | 0.0% |
| Rarely | 8.2% |
| Sometimes | 28.3% |
| Often | 34.7% |
| All the time | 28.8% |
|  |  |
| User experience level |  |
| Poor | 2.1% |
| Fair | 15.2% |
| Good | 39.1% |
| Very Good | 26.1% |
| Excellent | 17.3% |

|  |
| --- |
| **Table 2:** Fit indices of all models |
| **Model** | $$χ\_{SB}^{2}$$ | **df** | **RMSEA** | **CFI** | **SRMR** | **PSRMR** | **GFI** | **LSR** | **Path** |
| CFA-Expected | 0.16 | 199 | 0 | 1 | 0.046 | 0.052 | 0.99 | 0.40 | ExpE4-ExpE5 |
| CFA-Perform | 2.06 | 199 | 0 | 1 | 0.041 | 0.046 | 0.99 | 1.46 | PerfRs1\_PerfRs2 |
| Trait Convergent | 21.31 | 892 | 0 | 1 | 0.210 | 0.221 | 0.86 | 8.15 | PerfRl1\_PerfRl2 |
| Method Convergent | 1.94 | 901 | 0 | 1 | 0.056 | 0.059 | 0.98 | 7.38 | PerfRl1\_PerfRl2 |
| CU | 0.50 | 437 | 0 | 1 | 0.025 | 0.038 | 1.00 | 0.74 | PerfRl2\_ExpT3 |
| Fully Crossed CFA | Non-positive definite matrix |

All RMSEA 90% CIs were [0, 0], all p-values were close to 1. LSR = Largest Standardized Residual, Path = suggested error correlation for the corresponding LSR. When CFI values are greater than 1, they are automatically rounded off to 1 by software programs. When the chi-square values are less than the degrees of freedom, the RMSEA is set equal to 0 by the programs.

|  |
| --- |
| **Table 3: Factor and Structure Coefficients** |
|  | T |  | REL |  | RES |  | A |  | E |
|  | f/p | st |  | f/p | st |  | f/p | st |  | f/p | st |  | f/p | st |
| Exp\_T1 | 0.53 | 0.53 |  | 0.00 | 0.53 |  | 0.00 | 0.52 |  | 0.00 | 0.48 |  | 0.00 | 0.45 |
| Exp\_T2 | 0.48 | 0.48 |  | 0.00 | 0.48 |  | 0.00 | 0.48 |  | 0.00 | 0.44 |  | 0.00 | 0.40 |
| Exp\_T3 | 0.51 | 0.51 |  | 0.00 | 0.51 |  | 0.00 | 0.50 |  | 0.00 | 0.46 |  | 0.00 | 0.43 |
| Exp\_T4 | 0.56 | 0.56 |  | 0.00 | 0.56 |  | 0.00 | 0.55 |  | 0.00 | 0.51 |  | 0.00 | 0.47 |
| Exp\_Rl1 | 0.00 | 0.52 |  | 0.52 | 0.52 |  | 0.00 | 0.48 |  | 0.00 | 0.42 |  | 0.00 | 0.41 |
| Exp\_Rl2 | 0.00 | 0.54 |  | 0.54 | 0.54 |  | 0.00 | 0.50 |  | 0.00 | 0.44 |  | 0.00 | 0.42 |
| Exp\_Rl3 | 0.00 | 0.49 |  | 0.49 | 0.49 |  | 0.00 | 0.45 |  | 0.00 | 0.40 |  | 0.00 | 0.38 |
| Exp\_Rl4 | 0.00 | 0.51 |  | 0.51 | 0.51 |  | 0.00 | 0.47 |  | 0.00 | 0.41 |  | 0.00 | 0.40 |
| Exp\_Rl5 | 0.00 | 0.51 |  | 0.51 | 0.51 |  | 0.00 | 0.47 |  | 0.00 | 0.41 |  | 0.00 | 0.40 |
| Exp\_Rs1 | 0.00 | 0.48 |  | 0.00 | 0.44 |  | 0.48 | 0.48 |  | 0.00 | 0.39 |  | 0.00 | 0.40 |
| Exp\_Rs2 | 0.00 | 0.57 |  | 0.00 | 0.53 |  | 0.58 | 0.58 |  | 0.00 | 0.48 |  | 0.00 | 0.49 |
| Exp\_Rs3 | 0.00 | 0.59 |  | 0.00 | 0.55 |  | 0.60 | 0.60 |  | 0.00 | 0.49 |  | 0.00 | 0.50 |
| Exp\_Rs4 | 0.00 | 0.54 |  | 0.00 | 0.51 |  | 0.55 | 0.55 |  | 0.00 | 0.45 |  | 0.00 | 0.46 |
| Exp\_A1 | 0.00 | 0.50 |  | 0.00 | 0.45 |  | 0.00 | 0.45 |  | 0.55 | 0.55 |  | 0.00 | 0.48 |
| Exp\_A2 | 0.00 | 0.42 |  | 0.00 | 0.37 |  | 0.00 | 0.38 |  | 0.46 | 0.46 |  | 0.00 | 0.40 |
| Exp\_A3 | 0.00 | 0.49 |  | 0.00 | 0.44 |  | 0.00 | 0.44 |  | 0.54 | 0.54 |  | 0.00 | 0.48 |
| Exp\_A4 | 0.00 | 0.44 |  | 0.00 | 0.39 |  | 0.00 | 0.39 |  | 0.48 | 0.48 |  | 0.00 | 0.42 |
| Exp\_E1 | 0.00 | 0.50 |  | 0.00 | 0.47 |  | 0.00 | 0.50 |  | 0.00 | 0.53 |  | 0.60 | 0.60 |
| Exp\_E2 | 0.00 | 0.46 |  | 0.00 | 0.43 |  | 0.00 | 0.46 |  | 0.00 | 0.48 |  | 0.55 | 0.55 |
| Exp\_E3 | 0.00 | 0.51 |  | 0.00 | 0.48 |  | 0.00 | 0.51 |  | 0.00 | 0.54 |  | 0.61 | 0.61 |
| Exp\_E4 | 0.00 | 0.46 |  | 0.00 | 0.43 |  | 0.00 | 0.46 |  | 0.00 | 0.48 |  | 0.55 | 0.55 |
| Exp\_E5 | 0.00 | 0.55 |  | 0.00 | 0.51 |  | 0.00 | 0.55 |  | 0.00 | 0.58 |  | 0.66 | 0.66 |
| Perf\_T1 | 0.42 | 0.42 |  | 0.00 | 0.42 |  | 0.00 | 0.42 |  | 0.00 | 0.38 |  | 0.00 | 0.35 |
| Perf\_T2 | 0.36 | 0.36 |  | 0.00 | 0.36 |  | 0.00 | 0.36 |  | 0.00 | 0.33 |  | 0.00 | 0.30 |
| Perf\_T3 | 0.33 | 0.33 |  | 0.00 | 0.33 |  | 0.00 | 0.33 |  | 0.00 | 0.30 |  | 0.00 | 0.28 |
| Perf\_T4 | 0.43 | 0.43 |  | 0.00 | 0.43 |  | 0.00 | 0.43 |  | 0.00 | 0.39 |  | 0.00 | 0.36 |
| Perf\_Rl1 | 0.00 | 0.32 |  | 0.32 | 0.32 |  | 0.00 | 0.29 |  | 0.00 | 0.26 |  | 0.00 | 0.25 |
| Perf\_Rl2 | 0.00 | 0.41 |  | 0.41 | 0.41 |  | 0.00 | 0.38 |  | 0.00 | 0.33 |  | 0.00 | 0.32 |
| Perf\_Rl3 | 0.00 | 0.32 |  | 0.32 | 0.32 |  | 0.00 | 0.29 |  | 0.00 | 0.26 |  | 0.00 | 0.25 |
| Perf\_Rl4 | 0.00 | 0.28 |  | 0.28 | 0.28 |  | 0.00 | 0.26 |  | 0.00 | 0.23 |  | 0.00 | 0.22 |
| Perf\_Rl5 | 0.00 | 0.42 |  | 0.42 | 0.42 |  | 0.00 | 0.39 |  | 0.00 | 0.34 |  | 0.00 | 0.33 |
| Perf\_Rs1 | 0.00 | 0.31 |  | 0.00 | 0.29 |  | 0.31 | 0.31 |  | 0.00 | 0.25 |  | 0.00 | 0.26 |
| Perf\_Rs2 | 0.00 | 0.33 |  | 0.00 | 0.30 |  | 0.33 | 0.33 |  | 0.00 | 0.27 |  | 0.00 | 0.28 |
| Perf\_Rs3 | 0.00 | 0.55 |  | 0.00 | 0.52 |  | 0.56 | 0.56 |  | 0.00 | 0.46 |  | 0.00 | 0.47 |
| Perf\_Rs4 | 0.00 | 0.41 |  | 0.00 | 0.38 |  | 0.41 | 0.41 |  | 0.00 | 0.34 |  | 0.00 | 0.34 |
| Perf\_A1 | 0.00 | 0.38 |  | 0.00 | 0.34 |  | 0.00 | 0.34 |  | 0.42 | 0.42 |  | 0.00 | 0.37 |
| Perf\_A2 | 0.00 | 0.27 |  | 0.00 | 0.24 |  | 0.00 | 0.25 |  | 0.30 | 0.30 |  | 0.00 | 0.26 |
| Perf\_A3 | 0.00 | 0.35 |  | 0.00 | 0.32 |  | 0.00 | 0.32 |  | 0.39 | 0.39 |  | 0.00 | 0.34 |
| Perf\_A4 | 0.00 | 0.35 |  | 0.00 | 0.31 |  | 0.00 | 0.31 |  | 0.38 | 0.38 |  | 0.00 | 0.33 |
| Perf\_E1 | 0.00 | 0.37 |  | 0.00 | 0.34 |  | 0.00 | 0.37 |  | 0.00 | 0.39 |  | 0.44 | 0.44 |
| Perf\_E2 | 0.00 | 0.34 |  | 0.00 | 0.31 |  | 0.00 | 0.34 |  | 0.00 | 0.35 |  | 0.40 | 0.40 |
| Perf\_E3 | 0.00 | 0.31 |  | 0.00 | 0.29 |  | 0.00 | 0.31 |  | 0.00 | 0.33 |  | 0.37 | 0.37 |
| Perf\_E4 | 0.00 | 0.31 |  | 0.00 | 0.29 |  | 0.00 | 0.31 |  | 0.00 | 0.33 |  | 0.37 | 0.37 |
| Perf\_E5 | 0.00 | 0.39 |  | 0.00 | 0.36 |  | 0.00 | 0.39 |  | 0.00 | 0.40 |  | 0.46 | 0.46 |

T = Tangibles, REL = Reliability, RESP = Responsiveness, A = Assurance, E = Empathy; f/p = factor/pattern coefficient, st = structure coefficient

|  |
| --- |
| **Table 4: Factor Correlations** |
|  | Tangibles | Reliability | Responsiveness | Assurance | Empathy |
| Tangibles | 1 |  |  |  |  |
| Reliability | 1 | 1 |  |  |  |
| Responsiveness | 0.99 | 0.92 | 1 |  |  |
| Assurance | 0.91 | 0.81 | 0.82 | 1 |  |
| Empathy | 0.84 | 0.78 | 0.84 | 0.88 | 1 |



**Figure 1:** Confirmatory factor analysis model for the expected level of performance



**Figure 2:** Part of the correlated uniqueness model showing the error correlations between items measuring the same method