SERVICE QUALITY MEASURES: A TEST OF CONVERGENT VALIDITY AND TRAIT-METHOD EFFECTS

James M. Tolliver, The University of New Brunswick
Robert W. Armstrong, Murdoch University
Daniel F. Coleman, The University of New Brunswick

ABSTRACT

This study reviews and tests seven operationalizations of service quality to ascertain convergent validity. It compares these with earlier parallel operationalizations of job satisfaction, in order to assess the impact of the trait-method units common to both research domains. Results show that the service quality construct has limited convergent validity and further the construct does not explain a major portion of global service quality variance. Comparison of the service quality construct to the job satisfaction construct supplies some evidence of trait-method bias.

INTRODUCTION

"Convergent validity" requires that we have several different measures of the same construct (Campbell & Fiske, 1959). If a particular construct can only be measured by the use of a specific technique or with a single instrument, it raises the possibility that ensuing research results are traceable to the way the construct was measured, rather than to the construct itself. One example of such "method bound" research is the early work of Herzberg and his colleagues (e.g. Herzberg, Mausner, & Snyderman, 1959). This research led to a series of plausible but erroneous conclusions concerning the nature of job satisfaction and motivation. Before the results of this research programme were traced to the methodology employed the results were embraced by practitioners and utilized as the basis of perhaps thousands of company training programmes--programmes that were of doubtful value given the findings of subsequent research.

The requirement that a construct have convergent validity is consequently of interest in any field where researchers report the formulation of specific constructs or practitioners apply such constructs as the basis of management techniques. The area of service quality is no exception. If "service quality" can only be measured in one way or if various measures of service quality do not converge it raises the possibility that they are "method bound", misleading researchers and practitioners alike. In short, the construct and practice founded on it will be suspect, and subject to the same fate as Herzberg's early work. Fortunately, since 1988, when Parasuraman, Zeithaml and Berry published SERVQUAL, an early instrument designed to measure service quality, the service quality construct has been redefined and operationalized in a number of different ways. Consequently it is now possible to test the convergent validity of several "service quality" measures.

Campbell and Fisk (1959) also note that a construct is a "trait-method unit". Trait-method units are comprised of the stimuli which represent the construct, such as the questions asked on a questionnaire, **and** the methods employed to analyse the results (e.g. the mathematical methods used to determine a score on the questionnaire). It is also possible the trait-method unit employed to "tap" a particular construct determines the research results. In other words it is possible that any relationship discovered between a construct, such as service quality, and any other construct is due to an interaction between the questions asked and the way in which the answers to the questions are measured, scored, or analyzed. It is therefore useful to examine findings from different, but logically unrelated, research "domains" which employ the same trait-method units to determine if trait-method similarity has an effect on the reported results.

Simply put, any research using a construct such as service quality is suspect if 1) the construct can be measured in only one way (it lacks convergent validity) or 2) if the application of similar questions and similar mathematical techniques in a completely unrelated field gives **exactly** the same pattern of statistical results (indicating a possible trait-method artifact).

This paper consequently takes two approaches to the service quality question. We first evaluate the convergent validity of several measures that purport to measure service quality. This is done by

a traditional correlation analysis and a more appropriate hierarchical regression analysis conducted *within* the service quality domain. Second, we test for a trait-method effect.

The possibility of a trait-method effect is evaluated by comparing the pattern of results from service quality research to the pattern of results from job satisfaction research. Job satisfaction is an unrelated construct, but one which uses the same trait-method units as service quality research. This is a cross domain comparison.

The Trait-Method Units

Service quality and most job satisfaction instruments are comprised of the same trait-method units. In both traditions subjects often indicate, on a series of Likert scales, summary judgments of what SHOULD BE, what IS NOW, what they WOULD LIKE and the IMPORTANCE of a variety of items or "facets" concerning a job or a service (For clarity IS NOW refers to a job satisfaction construct; the same concept in the service quality domain is indicated by *Is Now*). Such responses are then combine into several dimensions or a general "index". This index is seen as indicative of general job satisfaction or service quality and is sometimes compared to a single global job satisfaction measure or global service quality measure (e.g. Cronin & Taylor, 1992). In some cases difference scores, (or "gap measures") are calculated by subtracting some perceived facet level from some desired level of that facet. Facet *Importance* is also often used as a weight, either for dimensions (e.g. Parasuraman, Zeithaml, & Berry, 1994), or for individual facets when summed as an index (Cronin & Taylor, 1994; Wanous & Lawler, 1972).

In sum, service quality measures and most job satisfaction measures use trait-method units consisting of the same question stems and the same mathematical formulations, often weighted difference scores. It is possible that using SHOULD BE, IS NOW, WOULD LIKE and IMPORTANCE stems and then using techniques such as "gap measures" to combine the responses to such stems, or by weighting such responses by IMPORTANCE, results in a specific *pattern* of research findings. This pattern might carry across research areas or "domains".

An examination of the impact of these specific trait-method units across domains has yet to be done, but it is certain that many of the trait-method units described above are problematic. Edwards (1991) identified a number of significant measurement and analytical problems that are traceable to these trait-method units in a body of literature he calls the "Person-Job", or "P-J Fit" literature. He concludes that the data collection and analytical methods currently used in the Person-Job literature, including job satisfaction, could well be misleading (Edwards, 1991). The literature Edwards specifically criticizes views work outcomes (such as job satisfaction) as a match between "desires" (such as needs, values, or preferences) and "supplies" (such as job attributes or workload). This is, in effect, the same analytical form ("gap" analysis) that is found in a lot of service quality research. Edwards (1991), however, confined his review to the P-J literature, including job satisfaction, leaving open the possibility that similar problems ***across*** domains--specifically in service quality--could be due to the use of such trait-method units.

The Job Satisfaction Construct. In 1972 Wanous and Lawler identified nine trait-method units that are prevalent in the job satisfaction literature. These trait-method units, presented in Table 1, still clearly define conceptualization and measurement in the job satisfaction field.

While the formulae presented in Table 1 are rather direct it is useful to briefly provide definitions of the terms used: MEAN OF FACET SATISFACTION is an average satisfaction score calculated by directly asking each subject "how satisfied are you with x?" (where x is a specific item or facet such as pay). IS NOW asks how much of a facet, such as pay, is present. It is a measure of goal attainment or need fulfilment (Wanous & Lawler, 1972). IMPORTANCE is a rating of a facet's importance. SHOULD BE is a "fairness" construct; specifically, it is the subjects' judgement of "how much" of a facet, such as pay, is "fair" or equitable given the input required of the job. Finally, WOULD LIKE is a measure of the desired or "ideal" levels of an outcome.

IS NOW, IMPORTANCE, SHOULD BE, and WOULD LIKE are usually rated on a series of

Table 1
Operational Definitions of Overall Satisfaction
Correlations with Single Item Measuring Overall Satisfaction[a]

Equation Numbers	Absolute Value Correlation (r)	Measure[b]
1	.61	Is Now
2	.60	Mean of Facet Satisfaction
3	.54	Would Like - Is Now
4	.50	Importance x Facet Satisfaction
5	.48	Importance x Is Now
6	.45	Importance x (Would Like - Is Now)
7	.44	Importance - Is Now
8	.39	Importance x (Should Be - Is Now)
9	.24	Should Be - Is Now

[a] From Wanous and Lawler (1972), Table 1, page 98. Our equation numbers do ***not*** parallel Wanous and Lawlers' since we have chosen to present their findings as a partial replication of this original table.

[b] This column contains a description of the measures in words. The mathematical representation of each measure can be found in Table 2.

seven point Likert scales that range from "a lot" to "a little" or, in the case of IMPORTANCE, from "very important" to "not very important". Subjects respond to these "stems" for specific facets such as pay, promotional opportunities, and relations with co-workers. Typically a job satisfaction questionnaire might ask a respondent how much pay there IS NOW or how much pay there SHOULD BE and also ask the subject to rate the IMPORTANCE of pay. While different job satisfaction questionnaires use different stems and different mathematical combinations of them, these different combinations can be classified as indicating overall satisfaction (equations 2 and 4, Table 1), indicating fulfilment (1 and 5, Table 1), indicating perceived equity (8 and 9, Table 1), and as desired or value models (3 and 6, Table 1) (Equation 7 is seen as "theoretically meaningless") (Wanous & Lawler, 1972; Evans, 1969).

Wanous and Lawler reasoned that if such a variety of different trait-method units were tapping a common construct, and the influence of this common construct was greater than the unique variance taped by the measures, the responses to these different trait-method units should be highly related. In other words, if a general "job satisfaction" factor was being tapped by these different measures they should be highly correlated. They tested the convergent validity of these nine trait-method units by using a twenty-three facet instrument that focused on such items as pay and supervision. The nine trait-method units represented by the equations in Table 1 failed the test of convergent validity. IS NOW (measure 1, Table 1) correlated most highly with a global satisfaction measure while the equity measure (measure 9, Table 1) achieved the lowest correlation. IMPORTANCE was not useful as a facet weight but was related to the impact a particular facet had on global satisfaction (Wanous

& Lawler, 1972). In other words these trait-method units had little in common and were not similarly related to a global measure of satisfaction. This raised the possibility that incompatible research findings were due to the use of measures that claimed to measure the same construct but in fact did not. And while not stated directly, a failure to find convergent validity raised the possibility that the job satisfaction construct could not be measured adequately at all.

By implication the same conclusions might be applied to service quality research since, as we show below, current service quality measures use exactly the same trait-method units which lacked convergent validity in the job satisfaction literature.

The Service Quality Construct. SERVQUAL (Parasuraman, Zeithaml, & Berry, 1988), is a twenty-two facet instrument that taps "service quality". The original SERVQUAL asked subjects to indicate what service SHOULD BE offered by a firm (denoted as "*Expectations*" (E)) and the *Perception (P)* that a target firm had the described facet. Service quality [Q], a gap measure, is defined as Q = P - E. The results are summed across facets to represent a total measure of service quality.

Subsequently Carman (1990) concluded that facet *Importance* (IMPORTANCE) may be more relevant than *Expectations* (SHOULD BE). Babakus and Boller (1992) questioned the usefulness of gap analysis, concluded that the SERVQUAL may be uni-dimensional, and tested *Importance* as a facet weight. They discovered *Importance* added little to the analysis.

By using a SHOULD BE stem and the perception (IS NOW) stem, Parasuraman, Zeithaml, and Berry, (1988), used a trait-method unit directly comparable to the "Equity" measure of job satisfaction (SHOULD BE - IS NOW, measure 9, Table 1) while Babakus and Boller, (1992) are, in effect, using those trait-method units denoted as 1, 5, 8, and 9 in Table 1. Babakus and Boller's findings (1992) parallel an earlier discovery in the job satisfaction literature that using IMPORTANCE as a weight is an ineffective technique (e.g. Blood, 1971; Evans, 1969). Similar to Wanous and Lawler's finding that IS NOW explained the most variance in a global measure of job satisfaction, SERVPERF (an IS NOW measure) explained more variation in a global measure of service quality than the SERVQUAL. Others (Teas, 1993) raise the issue of "Ideal Points" (a WOULD LIKE trait-method unit), and also question the dimensionality of the original and revised SERVQUAL models (Teas, 1994).

From this admittedly brief review it is clear that the trait-method units used in service quality research and job satisfaction research (as well as some other areas of the Person-Job fit literature) are the same. The extent of this similarity can be seen from Table 2 which pairs service quality trait-method units with comparable job satisfaction trait-method units.

The fact that the trait-method units used in service quality research are the same as those used in job satisfaction research presents an opportunity to evaluate the impact of the various trait-method units used. While convergent validity can be determined by comparing the results from using different service quality trait-methods--a within domain analysis--a cross-domain analysis can be conducted by comparing the pattern of results achieved using various service quality trait-method units to the results achieved with the same trait-method units used in job satisfaction. Consequently in the analysis below we have included gap measures which are still much in vogue (Oliver, 1997) and also included a measure of *Importance*. While we are aware of the current debate (e.g. Oliver, 1997) concerning the concepts of service quality and service satisfaction, we have included those trait-method units which various authors (e.g. Parasuraman, Zeithaml & Berry, 1988) have claimed to be service quality measures. This permits us to directly compare findings ***across*** domains and to ascertain if the limited impact of some variables, such as *Importance*, is due to an inappropriate analysis. Given differences in the domains, the samples, the cultures, the facets and the time lines (the data describe below were collected more than 20 years after the job satisfaction data used for comparison) it is likely that a similar result pattern is due to method variance traceable to the trait-method units employed rather than to similarities in the samples used. Thus, this paper addresses two research questions:

Table 2
Wanous and Lawler Job Satisfaction Operationalizations Compared with Possible Operationalizations of Service Quality

NUMBERS FROM TABLE 1	WANOUS/LAWLER DESCRIPTION	EQUATION	SERVICE QUAL. EQUIVALENT	EQUATION
1	Is Now	JS = Σ (Is Now)	Perceptions	$Q = \Sigma (P_i)$
3	Would Like - Is Now	JS = Σ (Would Like - Is Now)	Ideal - Perceptions	$Q = \Sigma (Id_i - P_i)$
5	Importance x Is Now	JS = Σ (Importance * Is Now)	Importance x Perceptions	$Q = \Sigma (I_i * P_i)$
6	Importance x (Would like - Is Now)	JS = Σ [Importance (Would Like - Is Now)]	Importance x (Ideal - Perceptions)	$Q = \Sigma I_i (Id_i - P_i)$
7	Importance - Is Now	JS = Σ (Importance - Is Now)	Importance - Perceptions	$Q = \Sigma (I_i - P_i)$
8	Importance x (Should Be - Is Now)	JS = Σ [Importance (Should Be - Is Now)]	Importance x (Expectations - Perceptions)	$Q = \Sigma I_i (E_i - P_i)$
9	Should Be - Is Now	JS = Σ (Should Be - Is Now)	Expectations - Perceptions	$Q = \Sigma (E_i - P_i)$

Research Question 1. Convergent Validity. What is the relationship between the different trait-method units of service quality and how are these operationalizations related to the global measure of service quality? If the measures of service quality have convergent validity we expect the various trait-method units to be highly correlated with each other and to be highly and uniformly correlated to a global measure of service quality.

Research Question 2. Trait-method effects. If the same trait-method units are used in different research domains are the results the same? Specifically, is the **pattern** of correlations between the trait-method units used in service quality research similar to the **pattern** of correlations between the trait-method units used in job satisfaction research? If there **is** a strong trait-method effect we would expect the pattern of correlations found between the service quality trait-method units to be the same as the pattern of relations between the job satisfaction trait-method units.

THE CURRENT STUDY

Sample

Customers and clients of over twenty different Australian organisations in fifteen different service industries were sampled. A total of 3000 questionnaires were mailed or hand delivered; 1135 subjects (38%) responded comprising a final sample of 18 different organisations across the fifteen different industries. In most cases a questionnaire was given to the nth person sitting in an airline terminal, restaurant, medical centre or bank queue. In other industries, random sampling methods taken from credit card listings were used. MBA students were also sampled via convenience method. In most cases, probability sampling methods were used. In some cases proprietors offered incentives to complete the surveys (e.g., discount coupons, free drinks, etc.). Consequently, the method of data collection was somewhat different for some of our subjects (i.e. mail or a hand delivered questionnaire). A test of the major variables, however, did not find any

difference due to the data collection method used.

The median number of responses within each industry is approximately 100. The respondents' average age is 29.81 years. There are 445 male and 620 female respondents; 70 persons did not indicate gender. The subjects are approximately evenly distributed across industries. The mean level of previous experience with all industries combined is 1.21 years with a minimum of zero (no experience) to a maximum 60 years.

Measurement Instrument

The original 22-item SERVQUAL scale developed to measure service *Perceptions* and *Expectations* was rewritten to include service *Importance* and *Ideal* service stems. The negatively worded questions (a problem in the SERVQUAL; Babakus and Boller, 1992) were rewritten in the positive. *Importance* was measured on each of the 22 Parasuraman, Zeithaml, and Berry, (1988) facets using a 7 point-scale (1 "Not Important"; 7 "Very Important"). *Ideal* service was measured for each facet by asking respondents to indicate "How much of an attribute should be present in an *Ideal* organization" (1 "None"; 7 "A Great Deal"). The *Ideal* scale differs from the *Expectations* scale in that the *Expectations* scale is based on the question, "Please show the extent to which you think XYZ should posses the feature described in each statement." In short, the *Ideal* scale asks respondents to imagine what they WOULD LIKE while the *Expectation* scale asks respondents to indicate what the service SHOULD BE for a particular firm. This wording attempts to directly tap the "ideal" point discussed in the literature (Parasuraman, Zeithaml, & Berry, 1994; Teas, 1993; Teas, 1994).

In addition to the 22 item SERVQUAL scale, a seven point (1 Low; 7 High) global measure of service quality, "Overall how would you rate the service quality of XYZ?" was included.

Analyses

Research Question 1. Convergent Validity. The relationship between the seven trait-method units used in service quality research may be analysed in two ways. The first method is to perform the calculations required by the various formulations contained in Table 2 then a) correlate these results with the global measure of service quality and b) determine the inter-correlations between the seven service quality operationalizations. This is the "analysis by correlation" method originally used by Wanous and Lawler (1972), recommended by Campbell and Fiske (1959), and used in a number of service quality studies. While it is a weaker analytic approach it is included for the purposes of comparison to earlier work.

The second, stronger, method uses hierarchical regression to enter the component parts of those trait-method units that employ difference scores separately -- a procedure which is more acceptable than using the correlation procedure. The reasoning for this "composite procedure" can be found in Schmidt and Wilson (1975), in Edwards and Cooper (1990) and in Evans (1991) and is explained more fully below.

RESULTS: QUESTION 1

Correlational Analyses

Table 3A contains the correlations of each service quality trait-method unit with the other service quality trait-method units. The correlation of the various service quality trait-method units with the global measure of service quality are compared in the last three columns of Table 3A. Table 3A requires that the sometimes "weighted" difference scores for each subject be summed across all 22 SERVQUAL items before the correlations are computed. While the content, number and stability of the dimensions which comprise the SERVQUAL has been the subject of debate (e.g. Carman, 1990), Cronin and Taylor (1994) contend that SERVQUAL is best represented by a single sum across all items.

The expectation of convergent validity is that measures be highly related to each other and to the global measure of service quality. While some trait-method units are related arguing for convergent validity, Table 3A shows that significant differences do exist. In particular, the relationships of the *Ideal* or *Expectations* trait-method units with other trait-method units are different. For example, the correlations between the *Ideal* (Id) trait-method units and *Perceptions*

Table 3A
Service Quality Equations Correlation Matrix[a]

Equation Number From Table 1		Service Quality Description	$\Sigma (P_i)$	$\Sigma (Id_i - P_i)$	$\Sigma (I_i * P_i)$	$\Sigma I_i (Id_i - P_i)$	$\Sigma (I_i - P_i)$	$\Sigma I_i (E_i - P_i)$	$\Sigma (E_i - P_i)$	Correlations with Global Measure[b]		
										This Study	C & T[c]	B & B[d]
1	$\Sigma (P_i)$	Perceptions	1	0.60	.90 (.91)	0.59	0.77	.82 (.80)	.82 (.81)	0.77	0.60	0.66
3	$\Sigma (Id_i - P_i)$	Ideal-Perceptions		1	0.42	0.99	0.66	0.64	0.64	0.48	--	--
5	$\Sigma (I_i * P_i)$	Importance x Perceptions			1	0.41	0.44	.57 (.63)	.55 (.66)	0.67	0.56	--
6	$\Sigma I_i (Id_i - P_i)$	Importance x (Ideal - Perceptions)				1	0.67	0.65	0.64	0.47	--	--
7	$\Sigma (I_i - P_i)$	Importance - Perceptions					1	0.91	0.92	0.64	--	--
8	$\Sigma I_i (E_i - P_i)$	Importance x (Expectations - Perceptions)						1	.99 (.98)	0.69	0.54	--
9	$\Sigma (E_i - P_i)$	Expectations - Perceptions							1	0.70	0.54	0.59

[a] All correlation coefficients were significant at the 0.0001 level.
[b] All correlations were significant at the 0.0001 level.
[c] Cronin and Taylor (1992).
[d] Babakus and Boller (1992).

(IS NOW) trait-method units is significantly smaller than the relations between the *Expectations* trait-method units and *Perceptions*. All such comparisons are statistically significant (p. = .001, or beyond, Fisher r to Z transformation). Another trend is that the *Ideal* trait-method units are closely related to each other, but are not as strongly related to the *Expectation* units. Finally, replicating previous research, all correlations remain relatively unchanged if *Importance* (I_i) is present or absent.

The correlations reported by Cronin and Taylor (1992)--who use this method--are reported in parentheses for comparison purposes. The relative magnitude of the correlations discovered in the Australian sample replicate the relative magnitude of those correlations reported by Cronin and Taylor (1992) with a United States sample (Kendall's Tau, λ = .83) indicating apparent cross-cultural stability.

Table 3A also reports correlations of the various measures with a single global measure of service quality. The correlations from Babakus and Boller (1992) and, again, Cronin and Taylor (1992) are included as comparisons. In Table 3A, the summation of the *Perceptions* ($Q = \Sigma(P_i)$) is most strongly related to a global measure of service quality (.77, p. < .001). The next highest correlation (.70) is the trait-method unit recommended by Parasuraman, Zeithaml, and Berry (1988) [$Q = \Sigma(E_i-P_i)$]. *Importance* (I_i) does not improve the correlation with the global measure. The *Ideal* (Id_i) trait-method unit does not improve the explained variance of the global measure (correlations .48 and .47). This pattern extends the findings of Cronin and Taylor (1992)

and Babakus and Boller (1992) concerning *Perceptions* ($Q = \Sigma(P_i)$).

Regression Results

While the "analysis by correlation" method has been used extensively, and was conducted in this paper for the purposes of comparison with earlier research, a more appropriate test of convergent validity is accomplished via hierarchical regression. As Teas (1994) notes the calculation of differences may result in sums that are at or near zero. This was the case for the *Should Be-Is Now, Would Like-Is Now and Importance-Is Now* data in this study and, as noted by one reviewer, further limits the usefulness of the "analysis by correlation" method. But the correlational approach is also questionable on theoretical rather than empirical grounds.

Schmidt and Wilson (1975) recommend hierarchical regression because the multiplication of two Likert scales can result in multiplying terms that contain deviations from "true" zero. For example, weighting *Perceptions* (P_i), by *Importance* (I_i) (to construct a trait-method unit such as equation 5 in Table 3A) is a multiplication of a score and its deviation from true zero (if any) with another score and its deviation from true zero. In Schmidt and Wilson's (1975) terms, the measure of *Perceptions* has a "true" component (P_i) and some unknown error component (e_p). The *Perceptions* measure is then actually $P_i + e_p$, and the *Importance* measure is similarly $I_i + e_I$. The multiplicative term of $I_i * P_i$ is actually $I_i P_i + I_i e_p + P_i e_I + e_I e_p$. Consequently, if a transformation is applied to I_i, to P_i or both, say by rescaling the data, this legitimate transformation could result in a substantial change in the correlation of $I_i * P_i$ with the global measure of service quality (or any other measure, even other composite operationalizations) (see Schmidt & Wilson, 1975). Since convergent validity is dependant on trait-method correlations, rescaling effects the conclusion of convergent validity or the lack of it.

Evans' (1991) suggestion is to consider any multiplicative form (such as many of the trait-method units in Table 1) as an interactive term in a linear regression, and to analyse any relationship involving such a term by first taking into account its component parts as main effects. This requires that those terms used for multiplication be first entered in a regression as "main effects" and that the impact of the multiplicative term subsequently be assessed. For example, in evaluating the relationship of $I_i,* P_i$ to another variable, the variables that comprise $I_i,* P_i$ (I_i, and P_i respectively) must be entered into a regression **first.** The test of any multiplicative trait-method unit's relationship with any other trait-method unit is then the ability of the interaction ($I_i * P_i$) to add significantly to the R^2 after the component parts, or "main effects" that comprise the interaction have been entered. This technique is preferred to the correlation approach although it "...relies on the assumption of linearity between the underlying psychological variables and the measures used." (Evans, 1991: 7).

Edwards and Cooper (1990) argue that terms which comprise difference (i.e. gap) measures, such as many of the trait method units in Table 1, must also be entered separately into the equation. This is so because **any** operationalization which uses a gap measure, even an absolute value, in the form of $A_i = a_0 + b_1 (E_i - P_i) + e$ is a more restrictive, misleading, form of the equation $A_i = a_0 + b_1 (E_i) - b_2(P_i) + e$ (For example, in the case of the SERVQUAL: A_i could be global satisfaction; a_0 the intercept, b's are conventional beta weights, P_i is the *Perceptions* measure and E_i is the *Expectations* measure. The error term is e.)

The restrictive model effectively **requires** the beta weights for both P_i and E_i to be the same and also **requires** that P_i be subtracted from E_i . When using the second model, however, these constraints **will** be met only if a "gap" measure is truly a good predictor of the dependant variable. The resulting equation **will** have beta weights approximately the same and the sign of the second beta weight **will** be negative due to the characteristics of the data. In the case of the more appropriate second model, however, the equation will arise from the data and not have been forced by the trait-method unit used. Consequently an equation such as 9 (Table 2) would be $Q = a_0 + b_1(E_i) - b_2(P_i) + e$.

To test the effect of these restrictions and to meet the objections outlined above we conducted two regressions using the trait-method units of service quality as independent variables and the global service quality measure as the dependant

Table 3B
Hierarchical Regression Results

Equation with Global Service Quality as dependent variable	R^2	Degrees of Freedom	F-Score
$\Sigma(P_i)$	0.60	919	1384.24
$\Sigma(I_i)$	0.60	919	521.45
$\Sigma(P_i)+\Sigma(Id_i - P_i)$	0.60	897	672.16
$\Sigma(P_i)+\Sigma(Id_i - P_i)+ \Sigma(I_i * P_i)$	0.60	865	434.56
$\Sigma(P_i)+\Sigma(Id_i - P_i)+ \Sigma(I_i * P_i)+ \Sigma I_i(Id_i-P_i)$	0.60	865	326.01
$\Sigma(P_i)+\Sigma(Id_i - P_i)+ \Sigma(I_i * P_i)+ \Sigma I_i(Id_i-P_i)+ \Sigma (I_i-P_i)$	0.60	865	266.09
$\Sigma(P_i)+\Sigma(Id_i - P_i)+ \Sigma(I_i * P_i)+ \Sigma I_i(Id_i-P_i)+ \Sigma (I_i-P_i)+\Sigma I_i(E_i-P_i)$	0.61	850	221.64
$\Sigma(P_i)+\Sigma(Id_i - P_i)+ \Sigma(I_i * P_i)+ \Sigma I_i(Id_i-P_i)+ \Sigma (I_i-P_i)+\Sigma I_i(E_i-P_i)+ \Sigma(E_i-P_i)$	0.61	850	190.67

Table 3C
Hierarchical Regression: Edwards and Cooper Method

Equation with Global Service Quality as dependent variable	R^2	Degrees of Freedom	F-Score
$\Sigma(P_i)$	0.60	919	1384.2
$\Sigma(I_i)$	0.60	919	521.45
$\Sigma(P_i)+\Sigma(E_i)$	0.60	883	687.33
$\Sigma(P_i)+\Sigma(E_i)+ \Sigma(I_i)$	0.60	870	451.97
$\Sigma(P_i)+\Sigma(E_i)+ \Sigma(I_i) +\Sigma(Id_i)$	0.60	865	437.56
$\Sigma(P_i)+\Sigma(E_i)+ \Sigma(I_i) +\Sigma(Id_i) +\Sigma(I_i * P_i)$	0.60	865	338.65
$\Sigma(P_i)+\Sigma(E_i)+ \Sigma(I_i) +\Sigma(Id_i) +\Sigma(I_i * P_i)+\Sigma(I_i * E_i)$	0.60	865	270.61
$\Sigma(P_i)+\Sigma(E_i)+\Sigma(I_i)+\Sigma(Id_i) + \Sigma(I_i * P_i)+\Sigma(I_i * E_i)+\Sigma(I_i * Id_i)$	0.60	850	220.02

Table 4
A Comparison of the Correlations from the Wanous and Lawler Equations and Service Quality Equations

Equation Number From Table I	Wanous and Lawler Equation	Service Quality Equivalent	Wanous & Lawler Correlation	Service Quality Correlation
1	JS = Σ (Is Now)	$Q = \Sigma(P_i)$	0.61	0.77
3	JS = Σ(Would Like - Is Now)	$Q = \Sigma(Id_i - P_i)$	0.54	0.48
5	JS = Σ(Importance * Is Now)	$Q = \Sigma(I_i * P_i)$	0.48	0.67
6	JS = Σ[Importance (Would Like - Is Now)]	$Q = \Sigma I_i(Id_i - P_i)$	0.45	0.47
7	JS = Σ(Importance - Is Now)	$Q = \Sigma (I_i - P_i)$	0.44	0.64
8	JS = Σ[Importance (Should be - Is Now)]	$Q = \Sigma I_i(E_i - P_i)$	0.39	0.69
9	JS = Σ (Should Be - Is Now)	$Q = \Sigma(E_i - P_i)$	0.24	0.70

variable. The global service quality measure was selected since service quality is seen to be a global concept. The results of these regressions are reported in Table 3B and Table 3C. Since in our analysis it is the strength of the relationship not its form that is of concern we report the R squares but not the beta weights.

Table 3B reports the results when the trait-method units, including the results of the gap measures, are regressed directly on global service quality. Since the components used to calculate the difference scores are **not** entered separately in Table 3B we can compare the results of the "gap" models as currently conceptualized to the results we obtain when the objections of Evans (1991) and Edwards and Cooper (1990) are met. The second regression utilized the components separately as recommended by Evans and others. Both regressions were conducted in stages for the reasons explained above.

The improvement of R^2 in table 3B shows that the summed *Perceptions* measure ($Q = \Sigma(P_i)$) is the best predictor of global service quality. The other composite operationalizations do not enter the regression. This means that the other trait-method units have nothing to add to explaining global service quality. This finding supports the correlation results (Table 2) and the research of Cronin and Taylor (1992).

It may have been possible that the results in Table 3B are due to the restrictions placed on the trait-method units by **first** calculating the difference scores and then utilising these scores in the regression, this is a likely interpretation given the problem outlined by Teas (1994) and noted earlier. Consequently Table 3C reports the results when the *components* of the composite operationalizations are entered separately (Evans, 1991; Edwards and Cooper, 1990).

In this case the simple *Perceptions* measure is again most closely related to the global measure of service quality.

All analyses--correlations, and both regressions--give similar results. The *Perceptions* measure is more closely related to global perceptions of service quality than any other trait method unit and other formulations add nothing to the analysis.

RESULTS: QUESTION 2

Research Question 2. Trait-method effects. The cross-domain effect is tested by an examination of the **pattern** of correlations between the trait-method units and the global measure of job satisfaction or service quality respectively. If there is a trait-method effect we would expect to see the pattern of correlations found in job satisfaction to be replicated in the service quality data. (The job satisfaction correlations are taken from Table 1 and the service quality correlations from Table 3A).

The correlations in Table 4 were compared using two methods. First a Spearman rank correlation test was run treating each trait-method unit as an "object" and the correlations associated with each trait-method unit as a "pair" of observations. If there was a series of trait-method effects those trait-method units that correlate highly with a global measure of job satisfaction would be those same trait-method units that correlate highly with the global measure of service quality. The effect within each trait-method unit was examined by transforming the correlations to Z scores and comparing them. These tests reveal that service quality operationalizations yield different coefficients in terms of both pattern and magnitude than the job satisfaction trait-method units used as comparisons. The rank correlation test was not significant, and, in 5 of the 7 cases, the sizes of the correlations were significantly different across research domains ($p = .001$).

The service quality *Expectations* (SHOULD BE) is a much better predictor of global service quality than SHOULD BE was at predicting global job satisfaction. In both equation sets, *Importance* (IMPORTANCE) did not improve the explained variance. *Perception* (IS NOW) was the best predictor of global service quality and global job satisfaction but the correlations are of different magnitude ($p = .001$).

DISCUSSION

This paper uses proposals by Edwards (1991), Evans (1991), and others to test the convergent validity of different trait-method units of service quality. It also compares the **pattern** of correlations between different service quality trait-method units and a global measure of service quality to the **pattern** of correlations between the same trait-method units used in job satisfaction research and a global job satisfaction measure (Wanous & Lawler, 1972). The purpose of this comparison is to determine if the same trait-method units result in a common pattern of results across two different domains--a possibility given the findings of Edwards (1991) and the objections of Edwards and Cooper (1990).

Findings show that some service quality trait-method units are weakly correlated to others. Further, some composite operationalizations are differentially correlated to global service quality with *Perceptions* (an IS NOW measure) most strongly related to the global measure of service quality. This is supported using hierarchical regression as well as the more traditional, but flawed, "analysis by correlation" method.

While some service quality trait-method units are related--making a reasonable argument for convergent validity --this result must be approached with caution. From the data it seems that the results are due to the commonality of a general perceptual "set" tapped by *Perceptions* (IS NOW). This might be encouraging for service quality researchers, yet the use of *Perceptions* as the trait-method unit of choice is a problem since much of the variance in the global measure of service quality still remains unexplained. This is significant since service quality is a summary judgement about overall excellence or superiority. It is also problematic because "Validity is represented in the agreement between two attempts to measure the same trait through maximally different methods" (Campbell & Fiske, 1959: 83). In short validation typically requires confirmation by measurement procedures that are independent and maximally different but related. If this can not be shown, the construct under study is

questionable on the grounds that it can, essentially, be measured in "only one way"--or, in the present study, effectively taped only by an IS NOW approach.

Finally, as others have noted, the concept of a number of trait-method units (e.g."fit" or gap measures; Edwards, 1991) are ubiquitous in psychology and related disciplines. In fact a large number of approaches in different research domains rest on a surprisingly small number of trait-method units such as those found in both the job satisfaction and service quality literatures just reviewed. In such instances it is useful to determine if the "findings" seen as unique in one domain are due to the phenomenon under study or due to the method of data collection and analysis, that is the trait-method unit, used. This paper attempts to provide such an answer in the field of service quality. This interpretation -- that patterns of correlations are being repeated across domains due to the trait-method unit used -- receives marginal support at best. The rank-order of correlations found by Wanous and Lawler (1972) between the trait-method units used and a global measure of job satisfaction is not repeated when the same trait-method units are used in service quality research. An important exception to the dissimilar results, however, is the fact that the *Perceptions* (IS NOW) facets, at the core of the responses as noted above, were more strongly correlated with the global measures of both service quality and satisfaction than any other single measure or composite operationalization of either concept. It could therefore be the case that there is a method bias in both areas--that the IS NOW perceptual set dominates perceptions both in the service quality and job satisfaction domains. This might further be the case if similar facets (or "roots") were examined. In this study the service quality facets used were, with the exception of rewording of negative items, taken from the SERVQUAL and were chosen to make this study comparable to other service quality work. The facets in the Wanous and Lawler (1972) work dealt not with quality issues but with items such as pay and promotion opportunities. To the degree the facets or "roots" are considered as part of the trait-method unit (e.g. Oliver, 1997) this approach reduces the possibility of significant findings since the facets used in the two domains are not common. Further the Wanous and Lawler (1972) work was conducted over 20 years earlier and in the United States. The current data is therefore separated from the Wanous and Lawler comparison data not only by culture (an Australian sample) but by a significant amount of time and, perhaps, social change. Consequently, finding any pattern similarities becomes more difficult and the discovery of a significant "IS NOW" effect raises the possibility that the trait-method problem is more severe than the current data would indicate.

Considered together these findings lead to the conclusion that there is some marginal evidence for a trait-method effect, and that convergent validity may be due to the presence of a single common "set" taped by primarily by an *Perceptions* (IS NOW) trait-method unit. This paper, however, does not graph the form of the "raw" service quality data generated against forms generated by different "fit" indices such as the IS NOW model. When Edwards (1991) conducted similar tests he concluded that the data generated could "...take on a variety of forms, ranging from a simple sloped plane to a complex curvilinear surface." (Edwards, 1991: 346). He further discovered that while different fit models might produce statistically significant results they could misrepresent the raw data form. A similar objection, but on very different grounds, is raised by Oliver (1997) and others with the observation that expectancy disconfirmation might be more adequately represented by an "s" curve. Our paper asked if misrepresentations across research domains are due to the trait-method unit used (of which linear analytic tradition is a part) but did not investigate the suitability of the *Perceptions* **form** within the Service Quality research domain. This we leave for another paper.

REFERENCES

Babakus, E. and G. W. Boller (1992), "An Empirical Assessment of the SERVQUAL Scale," *Journal of Business Research*, 24, 253-268.

Blood, M. R. (1971), "The Validity of Importance," *Journal of Applied Psychology*, 55, 487-488.

Campbell, D. T. and D. W. Fiske (1959), "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 56, (2), 81-105.

Carman, J. M. (1990), "Consumer Perceptions of Service Quality: An Assessment of the SERVQUAL Dimensions," *Journal of Retailing*, 66, (1), 33-55.

Cronin, J. J. Jr. and S. A. Taylor (1992), "Measuring Service Quality: A Reexamination and Extension," *Journal of Marketing*, 56, 55-68.

Cronin, J. J. and S. A. Taylor (1994), "SERVPERF Versus SERVQUAL: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality," *Journal of Marketing*, 5, 125-131.

Edwards, J. R. (1991), "Person-Job Fit: A Conceptual Integration, Literature Review, and Methodological Critique," *International Review of Industrial and Organizational Psychology*, C. L. Cooper and I. T. Robertson, (eds.) New York: Wiley, 6, 283-357.

Edwards, J. R. and Cooper, C. L. (1990), "The Person-Environment Fit Approach to Stress: Recurring Problems and Some Suggested Solutions," *Journal of Organizational Behaviour*, 11, 293-307.

Evans, M. G. (1969), "Conceptual and Operational Problems in the Measurement of Various Aspects of Job Satisfaction," *Journal of Applied Psychology*, 53, 93-101.

Evans, M. G. (1991), "The Problem of Analysing Multiplicative Composites: Interactions Revisited," *American Psychologist*, 46, (1), 6-15.

Herzberg, F., B. Mausner and Snyderman (1959), *The Motivation to Work*, New York: Wiley.

Parasuraman, A., V. A. Zeithaml L. L. and Berry (1988), "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perception of Service Quality," *Journal of Retailing*, 64, (1), 12-40.

Parasuraman, A., V. A. Zeithaml and L. L. Berry (1994), "Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for Further Research," *Journal of Marketing*, 58, 111-124.

Schmidt, F. L. and T. C. Wilson (1975), "Expectancy Value Models of Attitude Measurement: A Measurement Problem," *Journal of Marketing Research*, 12, 366-368.

Teas, R. K. (1994), "Expectations as a Comparison Standard in Measuring Service Quality: An Assessment of a Reassessment," *Journal of Marketing* 58, 132-139.

Teas, R. K. (1993), "Expectations, Performance Evaluation, and Consumers' Perceptions of Quality," *Journal of Marketing*, 57, 18-34.

Wanous, J. P. and E. E. Lawler, III (1972), "Measurement and Meaning of Job Satisfaction," *Journal of Applied Psychology*, 56, (2), 95-105.

Correspondence regarding this article should be sent to:
James M. Tolliver
The University of New Brunswick
Post Office Box 4400
Fredericton New Brunswick
CANADA E3B 5A3