

A psychometric assessment of an instrument to measure a service firm's customer-based corporate reputation

C. Boshoff*

Department of Business Management, University of Stellenbosch,
Stellenbosch 7600, Republic of South Africa
cboshoff@sun.ac.za

Received February 2009

Over the years, marketers have been cautioned repeatedly about the use of measuring instruments that do not demonstrate sufficient evidence of reliability and validity, particularly construct validity. The use of marketing instruments that do not demonstrate sufficient evidence of construct validity can lead to invalid results and erroneous conclusions. In short, construct validity is not just a “nice to have” in marketing research – it is at the heart of marketing decision-making.

This study assesses the psychometric properties of a recently published instrument by Walsh and Beatty (2007) to measure the customer-based corporate reputation of a service firm, and illustrates how, by using an approach proposed by Steenkamp and Trijp (1991) to assess the uni-dimensionality of the underlying dimensions of a multi-dimensional construct, one can enhance the construct validity of such an instrument. The result is a shortened, 17-item instrument with superior psychometric properties compared to the longer 31-item instrument.

*To whom all correspondence should be addressed.

Introduction

References to the reputation of both individuals and business firms seem to have increased in recent times. There are many celebrated cases of business activities that have harmed companies' reputations. Some are more fortunate, and suffer the indignation of defending a once-proud reputation only very occasionally. The Hewlett-Packard spying scandal in 2006 (Darlin, 2006) is a case in point.

The reputation of other companies seems to be under constant scrutiny. Walmart is regularly accused of driving US jobs abroad and trampling workers at home (Collins, 2004). “McDonalds sells unhealthy food” is a refrain that the company has had to deal with more than once. Microsoft is often reminded about their undesirable market dominance and glitches in their software (Alsop, 2004). In some cases reputational damage can be fatal or close to fatal. Enron in the USA and Parmalat in Europe are well-documented examples.

Reputational damage and threats to reputations raise several questions. These include: Can a corporate reputation be measured? If so, what are the underlying dimensions of a corporate reputation?

Against this background Walsh and Beatty (2007) have proposed an instrument to measure the corporate reputation of a service firm. The primary purpose of this study is to assess the psychometric properties and in particular the construct validity of this scale as marketers have been cautioned repeatedly about the use of measuring instruments

that do not demonstrate sufficient construct validity. The use of marketing instruments that do not demonstrate sufficient evidence of construct validity can lead to invalid results and erroneous conclusions. Also, the validity of constructs is a necessary condition for theory development and testing and, therefore, construct validity lies at the very heart of scientific progress in marketing.

The analyses reported here illustrates how, by using an approach proposed by Steenkamp and Trijp (1991) to assess the uni-dimensionality of the underlying dimensions of a multi-dimensional construct, one can enhance the construct validity of such an instrument.

This practical illustration is preceded by a review of the concept corporate reputation against the background of sound measurement principles and in particular the role that Structural Equation Modelling (SEM) can play in this process. The study concludes with a discussion and suggestions of future research.

The concept “corporate reputation”

There is not full consensus in the literature as to what constitutes a corporate reputation. In addition, the concepts “corporate image” and “corporate reputation” are often confused. Brown, Dacin, Pratt and Whetton (2006) argue that corporate image is of a short-term nature, and is what the firm wants others to believe about them – a sort of ideal state that the firm tries to project. Corporate reputation, on the other hand, is a longer-term predisposition towards a firm, and is influenced by the views of other stakeholders

and, as a result, is not entirely manageable by the firm. A reputation can also differ from one stakeholder group to another. An investor disappointed by recent returns on his/her investment in the shares of a firm may hold a very different point of view about that firm from that of a satisfied, loyal customer.

Despite the apparent confusion on a definition of the concept, there appears to be a fair amount of agreement on the following aspects of the concept. A corporate reputation:

- is intangible
- is a collective term
- is based on past actions and interactions
- is based on perceptions; and
- is multi-dimensional.

More importantly, corporate reputation is an "...attitude-like evaluating judgment of a firm".

Based on these principles, Walsh and Beatty (2007:129) define the corporate reputation of a service firm as "... the customer's overall evaluation of a firm based on his or her reaction to the firm's goods, services, communication activities, interactions with the firm and/or its representatives or constituencies (such as employees, management or other customers) and/or known corporate activities."

Given the fact that services are characterised primarily by its intangibility (Murray, 1991; Murray & Schlacter, 1990) some argue that for a service firm a favourable reputation is even more important than for those marketing physical entities. "We're not like Coca-Cola where people buy the product off the shelf or in a vending machine .." says Joan Lollar of FedEx (Alsop, 2004:5) when discussing the importance of this service firm's reputation.

A firm's reputation can have both positive and negative outcomes. The benefits of a favourable corporate reputation have been well documented. These benefits include higher levels of positive word-of-mouth and even the luxury to charge a price premium. Other benefits usually cited when describing the advantages of a positive reputation include high levels of trust among customers, lower risk perceptions, and higher entry barriers for potential competitors. The potential disadvantages of a positive reputation are seldom discussed in the literature, but one of them could be that such firms have to deal with customer expectations that are substantially higher than those of firms whose reputation is not at the same level.

Measuring corporate reputation

Attempts to properly operationalise the concept "corporate reputation," to identify its underlying dimensions, and to develop a measuring instrument with psychometric properties beyond reproach have been limited, to say the least. Walsh and Beatty (2007) tried to address this gap with a study based on the earlier work of Fombrun, Gardberg and Sever (2000). The end result was 31-item instrument capturing five dimensions of what they believe should constitute a corporate image, namely: *Customer orientation,*

Reliable and financially strong, Product and service quality, Social and environmental responsibility and Employer relations.

A close look at the results reported by Walsh and Beatty (2007), and particularly the fit indices of their confirmatory factor analysis model, suggest that their model has certain limitations. More disconcertingly, those limitations seem to have manifested themselves again during their attempt to validate their model in a cross-validation sample.

This dilemma brings into focus the question of scale development procedures and the optimal utilisations of the most contemporary techniques in scale development, to ensure that marketers use instruments with sufficient evidence of construct validity in their research endeavours.

Scale development

In the late 1970s, Gilbert Churchill (1979) proposed a well-accepted procedure for the development of valid and reliable multi-item instruments. This process consists of six steps: domain specification, generation of questionnaire items, empirical surveying, an iterative process of scale purification based on reliability assessment and validity checks and, finally, the development of norms.

Since then, scale development and scale validation have received some attention in the marketing literature. These include contributions by Hinkin (1995), Malholtra, Peterson and Kleise (1999), Rossiter (2002) and Diamantopolous (2005).

A central theme in all of them is the concept of construct validity. Valid measurement is, according to Peter (1979), the *sine qua non* of science. He even argues that if a discipline does not use instruments that are valid and reliable, it cannot be regarded as a science. At the most basic level, construct validity means that the set of items making up a measuring instrument must measure only one thing in common (Hattie, 1985) and refers to the degree to which a construct achieves empirical and theoretical meaning (Bagozzi, 1980; Peter, 1981). In marketing terms, this means that the underlying items making up a scale are alternative indicators of the same construct (Gerbing & Anderson, 1996)

The ability of marketing researchers to assess the construct validity of the measuring instruments used in research has been enhanced in recent years with the availability of statistical procedures such as confirmatory factor analysis, using a structural equation modelling approach, which provides additional evidence of construct validity (Gerbing & Anderson, 1996; Tull & Hawkins, 1993), something that has not been available until relatively recently.

The contribution of structural equation modelling

Thanks to the pioneering work of statisticians such as Jöreskog, Sörbom, Van Thillio, Keesling, Wiley, Browne, and more recently social scientists like Bagozzi, Bentler,

Mulaik, and Huyduk, we today have at our disposal a technique known as “structural equation modelling.”

This technique has enabled researchers, for the first time, to explain a structure or a pattern, or a series of inter-related dependence relationships simultaneously among a set of latent constructs (unobserved, non-measurable or theoretical), each measured by one or more manifest (observed, empirical or measurable) variables that are measures or indicators (usually questionnaire items) of the unobservable latent variable, and has allowed researchers to objectively compare a theoretical model with empirical data.

Structural equation modelling has two important characteristics: the estimation of multiple interrelated dependence relationships, and the ability to represent unobserved concepts in these relationships, while accounting for measurement error in the estimation process. The new technique's greatest value lies in its potential to improve theory development and testing (Mackenzie, 2001).

The advantages of structural equation modelling¹

Several advantages of using SEM have been described in the literature. Two advantages are of particular benefit to marketing researchers in general and scale developers in particular. These advantages are the ability to control for measurement error, and to provide statistical tests of construct dimensionality.

Before the introduction of SEM, researchers had no way of dealing with measurement error, but SEM methods are able to take measurement error into account. In other words, SEM, unlike regression analysis, for example, allows a researcher to estimate the relationships among constructs that are free from measurement error. This benefit is important, because most measures used in marketing and consumer research reflect not only the construct they are intended to represent, but also random and systematic measurement error.

Another major advantage of SEM is that it can improve scale development by providing statistical tests of construct dimensionality (MacKenzie, 2001). The technique also offers new indices of construct/item reliability that are more accurate than traditional methods, as well as more rigorous procedures for evaluating discriminant validity. When cross-validating a measuring scale, SEM offers the researcher a powerful tool to also assess convergent validity. In addition, by using SEM, one can assess the stability of a measuring instrument by placing increasingly stringent restrictions on the model as the validation process proceeds (Steenkamp & Van Trijp, 1991).

Construct validity

The validity of constructs is a necessary condition for theory development and testing, and therefore, construct validity lies at the very heart of scientific progress in marketing research. “Construct validity” refers to the degree to which a construct achieves empirical and theoretical meaning (Bagozzi, 1980; Peter, 1981). In the literature, it has been proposed by several authors (see Nunnally & Bernstein, 1994; Churchill, 1979; Bagozzi, 1980; Judd, Jessor & Donovan, 1986; Gerbing & Anderson, 1996) that the following criteria should be met before any construct validity claims can be made:

- (1) uni-dimensionality,
- (2) within-method convergent validity,
- (3) reliability,
- (4) stability,
- (5) across-method convergent validity and discriminant validity, and
- (6) nomological validity.

Many researchers still use rather straightforward techniques such as coefficient alpha, exploratory factor analysis, and bi-variate correlations to assess the criteria for construct validity. While these traditional techniques are valuable, the emergence of covariance structure models and the widespread availability of accompanying computer programmes such as LISREL (Jöreskog & Sörbom, 1988) provide the researcher with a powerful new tool for more detailed assessment and refinement of the construct validity of marketing measurement instruments. Basically, there are two primary advantages of SEM over the traditional methods, pertaining to the construct's empirical and theoretical meaning. Firstly, it provides a test of the theoretical structure of the measurement instrument, i.e., the relationship of the construct to its measures. Secondly, the relationships between the construct and other constructs can be tested without the bias that measurement error introduces. Both advantages are relevant for theory building in marketing science as well as in applied settings where unbiased estimates of the measure's reliability, stability, and validity are also of importance.

Objectives

The primary objective of this study was to subject to psychometric testing an instrument proposed by Walsh and Beatty (2007) to measure the customer-based corporate reputation of a service firm. During its development and validation the authors of this instrument reported, amongst others, fit indices that can at best be described as precarious. In fact, when trying to validate their scale using a validation sample, the instrument fared worse than expected and a further three items had to be deleted to achieve a barely acceptable fit. This manoeuvring raises some serious concerns about the construct validity and in particular, the uni-dimensionality of the instrument.

¹This section is based extensively on: Mackenzie, S.B. 2001. Opportunities for improving consumer research through Latent Variable Structural Equation Modelling, *Journal of Consumer Research*, 28(1): 159-166.

Methodology

Data collection

In this study, the 31-item instrument proposed by Walsh and Beatty (2007) to measure a service firm's customer-based corporate reputation was used to collect the data. Each statement was linked to a 7-point Likert-type scale where a 7 was anchored by *Strongly Agree* and 1 by *Strongly Disagree*. Each respondent was asked to evaluate the commercial bank they used most often.

Sampling

The instrument to measure a service firm's customer-based corporate reputation proposed by Walsh and Beatty (2007) was administered to a convenience sample of 340 undergraduate commerce students. Two qualifications were posed: respondents had to be at least 20 years old, and had to be a client of a commercial bank (had to have an existing savings or cheque account). This sample is labelled as sample 1.

The cross-validation sample consisted of a randomly selected national sample of 167 adults who were interviewed by telephone by a commercial research company, using one of their data bases as sampling frame. As was the case with sample 1, two qualification questions ensured that all respondents were at least 20 years of age and were existing clients of a commercial bank.

Data analyses procedures

The data from a sample of 340 students (referred to as sample 1) who completed the 31-item instrument proposed by Walsh and Beatty (2007) to measure the customer-based reputation of a service firm were subjected to an exploratory factor analysis to assess the underlying dimensionality of the data. Based on the results of the exploratory factor analysis, some items were deleted from the data set.

The remaining items were then subjected to a Robust Maximum Likelihood Confirmatory Factor Analysis using LISREL 8.80. The resultant output was inspected for admissibility (particularly the completely standardised solution) and for sources of poor fit, including the modification indices.

Once we had a revised instrument demonstrating sufficient evidence of uni-dimensionality for each underlying dimension, we cross-validated the new instrument using a randomly selected national sample of 167 adults.

Empirical results

Exploratory factor analysis

To assess the discriminant validity of the items measuring corporate reputation as proposed by Walsh and Beatty (2007), an exploratory factor analysis was conducted. In line with Walsh and Beatty's (2007: 134) reasoning, we conducted a principal axis factor analysis with an oblique

rotation, the latter because they anticipated the underlying factors to be correlated.

The KMO measure of sampling adequacy was 0,899, and Bartlett's Test of Sphericity proved to be significant ($p < 0,001$) suggesting that the data were factorisable.

The exploratory analysis clearly demonstrated that the factor structure was consistent with the five factors (explaining 66.5% of the variation) reported by Walsh and Beatty (2007) but that several items did not demonstrate sufficient discriminant validity. Seven items either loaded on more than one factor or did not load to a sufficient extent (loadings ≥ 0.40 were regarded as statistically significant) on any factor. Deleting these items in an iterative process resulted in a five-factor solution consisting of 24 items (see Table 1).

Table 1 also shows that the respondents seem to have regarded item REL 2 (XXX Bank seems to recognise and take advantage of market opportunities) as a measure of the *Product and service quality dimension*.

The next step was to assess the construct validity five-factor structure of the 24-item instrument by means of a confirmatory factor analysis (CFA).

Sample 1: Multivariate normality assessment

As the choice of an estimation method in structural equation modelling is influenced by distributional properties of the data, the first step in the confirmatory factor analysis phase of the study was to assess the multivariate normality of the data. The null hypothesis considered was:

H^0 The data in sample1 demonstrate sufficient evidence of multivariate normality

To assess the multivariate normality of the data (skewness and kurtosis), LISREL 8,80 was used. The test result (Skewness and Kurtosis $\chi^2 = 1338,26$ $p < 0,001$) revealed that the assumption of multivariate normality did not hold for this data set, suggesting that the null hypothesis (H_0) had to be rejected. The result was that in all subsequent analyses, the Robust Maximum Likelihood (RML) estimation method was used, rather than the more conventional Maximum Likelihood (ML).

Sample 1: Confirmatory factor analyses results

The first step in the data analysis process was to assess (or to confirm) the five-factor structure for consumer-based corporate reputation in sample 1 using the original 31-item instrument proposed by Walsh and Beatty (2007). The confirmatory factor analysis model fit statistics (see Table 2) of the Robust Maximum Likelihood analysis ($\chi^2 = 826,10$ ($p < 0,00$), χ^2/df ratio = 1,948; RMSEA = 0,0613; ECVI = 3,850) suggested that the measurement model for the Walsh and Beatty (2007) customer-based corporate reputation instrument did not fit the data particularly well. This poor fit may be due to problems related to the uni-dimensionality of the underlying dimensions of corporate reputation.

Table 1: Rotated factor loadings

	Factor				
	1 Employee relations	2 Reliable and strong company	3 Social responsibility	4 Customer orientation	5 Quality products and services
EMPLOY1	0,754				
EMPLOY2	0,710				
EMPLOY3	0,611				
EMPLOY7	0,594				
EMPLOY6	0,594				
EMPLOY5	0,587				
EMPLOY4	0,412				
REL6		0,890			
REL7		0,845			
REL5		0,585			
REL8		0,517			
REL4		0,483			
SOCIAL2			0,893		
SOCIAL3			0,846		
SOCIAL1			0,562		
SOCIAL4			0,448		
CUST3				0,800	
CUST2				0,797	
CUST1				0,745	
CUST4				0,527	
QUAL5					0,601
QUAL4					0,547
QUAL3					0,506
REL2					0,408

The seven items that had to removed due to poor discriminant validity were the following:

Dimension: Customer orientation

XXX Bank takes customer rights seriously (CUST 5)

XXX Bank seems to care about all of its customers regardless of how much money they spend with them (CUST 6)

Dimension: Reliable and financially strong company

XXX Bank tends to outperform competitors (REL 1)

XXX Bank looks like it has strong prospects for future growth (REL3)

XXX Bank appears to be aware of its responsibility to society (REL9)

Dimension: Product and service quality

XXX Bank offers high quality products and services (QUAL1)

XXX Bank is a strong, reliable company (QUAL2)

Table 2: Fit indices: sample 1

	Original Walsh and Beatty instrument used in this study (31 items)	Revised model based on EFA results (24 items)	Revised model based on Steenkamp and van Trijp (17 items)
Sattora-Bentler χ^2	826,10	490,89	175,18
df	424	242	109
χ^2 / df	1,948	2,02	1,601
RMSEA	0,0613	0,0639	0,0491
% confidence interval for RMSEA	0,0551 - 0,0675	0,0557 – 0,0720	0,0352 - 0,0622
ECVI	3,850	2,408	1,044

Uni-dimensionality: Sample 1

To assess the uni-dimensionality of an instrument, Steenkamp and Trijp (1991) recommend that in the case of poor model fit, one should assess the standardised residuals for the potential reasons. Although the measurement model fit the data reasonably well (RMSEA = 0.0613), we nevertheless considered the standardised residuals. An inspection of the standardised residuals did reveal several values higher than the 2.58 cut-off value proposed by Jöreskog and Sörbom (1988), which suggested that some misspecification might have occurred, raising concerns about the uni-dimensionality of the reputation instrument.

As a result of this assessment a further seven items were removed from the instrument. These seven items were:

- CUST4 XXX Bank treats its customer fairly
- EMPLOY1 XXX Bank looks like a good company to work for
- EMPLOY5 XXX Bank seems to have good employees
- EMPLOY6 XXX Bank seems to maintain high standards
- REL6 XXX Bank has strong record of profitability
- REL7 XXX Bank is doing well financially
- SOCIAL4 XXX Bank appears to support good causes

After the removal of the said seven items it was decided to re-run the CFA model using only the remaining 17 items to measure the five dimensions of corporate reputation.

The improved model fit statistics reported in Table 2 ($\chi^2 = 175,18$ ($p < 0,00$), χ^2/df ratio = 1,601; RMSEA = 0,0491; ECVI = 1,044) suggest that the removal of the said seven items was well justified. In other words, the revised and shortened 17-item instrument to measure customer-based corporate reputation demonstrated sufficient evidence of uni-dimensionality.

Cross-validation: Sample 2

When developing a new scale, cross-validation is desirable because there is always the possibility that one has capitalised on chance. The 31-item customer-based corporate reputation instrument was again administered to a second, cross-validation national sample of 167 respondents (clients of several banks) similar to the previous sample. The offending seven items with the high standardised residuals in the first sample were also included, and the *a priori* proposition that they needed to be removed was again empirically confirmed. The model fit statistics for the 24-item instrument were ($\chi^2 = 432,78$ ($p < 0,00$), χ^2/df ratio = 1,780; RMSEA = 0,0689; ECVI = 3,306), very similar to the first sample (Table 3). An inspection of the standardised residuals confirmed that the removal of the offending seven items would be well justified. The confirmatory factor

analysis model was thus re-run without the seven offending items (thus only 17 items) resulting in a substantial improvement in model fit ($\chi^2 = 175,18$ ($p < 0,00$), χ^2/df ratio = 1,601; RMSEA = 0,0548; ECVI = 1,044) as shown in Table 3.

Table 3: Fit indices: cross validation sample

	Original Walsh and Beatty instrument used in this study (31 items)	Revised model based on EFA results (24 items)	Revised model based on Steenkamp and van Trijp (17 items)
Sattora-Bentler χ^2	727,84	432,78	147,86
df	424	242	109
χ^2/df	1,720	1,78	1,36
RMSEA	0,0657	0,0689	0,0463
% confidence interval for RMSEA	0,0575 – 0,0737	0,0583 – 0,0793	0,0247 – 0,0643
ECVI	5,252	3,306	1,421

The results of the confirmatory factor analyses for the cross-validation sample are summarised in Table 3, and demonstrate the consistently better fit of the “shortened” 17-item instrument.

Table 4 provides a comparison of the fit indices reported by the developers of the original instrument (first two lines of Table 4) to measure customer-based corporate reputation, as well as those that emanated from this study. The table reveals the consistently superior fit of the 17-item instrument over the original 31-item instrument.

The reliability and validity of the shortened version of the corporate reputation instrument

Despite our argument that the shortened version of the customer-based corporate reputation instrument is superior to the longer version proposed by Walsh and Beatty (2007), we still need to provide evidence of its reliability and within-method convergent validity.

Reliability

Reliability is generally regarded as a necessary condition for validity (Peter, 1979). Although not unique to structural equation modelling, the reliability of a proposed instrument should therefore also be assessed. Both the Cronbach alpha coefficients for the individual samples and the composite liabilities of the underlying dimensions were above the generally accepted cut-off value of 0.7 (Peter, 1979) in both samples.

Table 4: Comparison of fit indices⁽¹⁾

MODEL	FIT INDICES				
	χ^2	χ^2/df	RMSEA	GFI ⁽²⁾	AGFI ⁽²⁾
Walsh and Beatty Sample 1	1 128,63	3,43	0,0800	0,90	0,90
Walsh and Beatty CV sample ⁽³⁾	Not reported	2,90	0,0800	0,92	0,91
This study sample 1: 31 items	826,10	1,948	0,0613	----	----
This study sample 1: 17 items	175,18	1,601	0,0491	----	----
This study CV: 31 items	727,84	1,720	0,0657	----	----
This study CV: 17 items	163,35	1,499	0,0548	----	----

1) CV = cross-validation sample

2) The reporting of the GFI and AGFI is inappropriate when the RML estimation method is used

3) After the removal of a further 3 items

Table 5: Reliability assessment

	Internal consistency Sample 1: Students	Internal consistency Sample 2: Cross-validation sample	Composite reliability (Total sample)
Customer orientation	0,85	0,88	0,86
Employer	0,78	0,88	0,84
Reliability	0,78	0,84	0,84
Quality	0,79	0,84	0,81
Social	0,83	0,83	0,85

Within-method convergent validity

Several ways to assess the within-method convergent validity of an instrument have been proposed. The statistical significance of the regression coefficients, the correlation of the item with the construct, and overall fit of the model are all indicators of within-method convergent validity. In this model (the 17-item instrument), all regression coefficients are strongly significant ($p < 0.000$), with the lowest t-value being 8.669 in sample 1 and 6.980 in sample 2, the cross-validation sample. All items correlate significantly with each underlying dimension, while the overall model fit has already been alluded to. All these measures point to the excellent within-method convergent validity of the shortened 17-item corporate reputation instrument.

Table 6: Average variance extracted per dimension

	Average variance extracted (AVE)
Customer orientation	0,67
Employer relations	0,57
Reliable and strong company	0,57
Quality of products and services	0,59
Social and environmental responsibility	0,65

Further evidence of the convergent validity of the shortened 17-item version of Walsh and Beatty's corporate reputation scale is the fact that the average variance extracted (AVE) of each dimension is above the 0.50 threshold suggested by Bagozzi and Yi (1988).

CONCLUSION

Scale development to ensure the reliability and validity of measuring instruments has made some significant advances in recent years. Structural equation modelling techniques have made it possible to develop measuring instruments for use in marketing research that are truly valid and reliable.

The results reported by Walsh and Beatty (2007) in developing an instrument to measure the customer-based corporate reputation of a service firm are a cause for concern. After what appears to be an exhaustive scale purification process, fit indices that are less than impressive were reported (e.g. RMSEA of 0.08). When they attempted cross-validate the instrument, the fit indices appear to have been even worse, and a further three items had to be removed to attain what can at best be described as a reasonable fit (RMSEA of 0.08).

In this study, using a different scale development and assessment approach to assess the construct validity and particularly the uni-dimensionality of the instrument proposed to measure the customer-based corporate reputation (31-items), it appeared as if using a shortened version of the instrument (17-items) demonstrates superior levels of uni-dimensionality and convergent validity and thus construct validity.

A number of reasons for the limitations of the longer version can be identified. The first reason is the operationalisation of the dimensions of the construct corporate reputation. For instance, the dimension of *Product and service quality* appears to be problematic for measuring the reputation of a service firm. Some of the items used as measures of *Product and service quality* seem to refer to products rather than services. Other items refer to innovation (develops innovative services) and value for money (offers services that are good value for money), which are difficult to justify as measures of service quality.

It appears that the longer version of Walsh and Beatty's (2007) corporate reputation instrument may have suffered from response error to some extent, hence the rather poor fit indices they report.

Other items that are problematic purely from a face-validity perspective are:

- “XXX service firm appears to be aware of its responsibility to society” as a measure of the *Reliable and financially strong company* dimension. One could argue, from a purely face-validity perspective, that this item could be a measure of the *Social and environmental responsibility* dimension of corporate reputation.
- “XXX service firm is a strong and reliable company” as a measure of the *Product and service quality* dimension. From a face-validity perspective, this item appears to be an item measuring the *Reliable and financially strong company* dimension of corporate reputation.
- “XXX service firm stands behind the services that it offers” as a measure of the *Product and service quality* dimension rather than the *Reliable and financially strong company* dimension of corporate reputation.
- Could “making an effort to create jobs” be regarded as a measure of social responsibility of a service firm operating in the private sector?

The fit of these items seems to suggest that the operationalisation of the underlying dimensions of the construct corporate reputation could be questioned. If these items had been assessed for uni-dimensionality as suggested Steenkamp and Trijp (1991), many of the problems highlighted above may have been avoided.

The poor fit indices reported by Walsh and Beatty's (2007) can possibly be attributed to the data analysis option they chose. Studies by both Satorra and Bentler (1988 and 1994) and Hoogland and Boomsma (1998) have shown that when the assumption of multivariate normality is violated, the use of Maximum Likelihood as an estimation method in structural equation modeling is inappropriate. Unfortunately, the Walsh and Beatty (2007) paper does not report whether or not they assessed the distributional properties of their data. If they erroneously assumed that the distribution of their data was multivariate normal and used Maximum Likelihood (ML) as estimation method rather than the Robust Maximum Likelihood (RML) proposed for data where the assumption of multivariate normality is violated. This decision may have effected their reported fit indices. In this study, we assessed the distributional properties of our data. The resultant conclusion was that Robust Maximum Likelihood was the most appropriate estimation to use under the circumstances.

However, the problems with Walsh and Beatty's (2007) corporate reputation instrument identified here may have occurred much earlier. Early in the scale purification process, they chose to assess the discriminant validity of an early version of the instrument, using an exploratory factor analysis. They used a principal-axis rotation, and argued that they expected the underlying factors to be correlated. This is a dubious argument. To assume that a service firm that is customer-oriented is also environmentally responsible, is questionable. Similarly, why would employer relations be related to product and service quality?

Limitations and future research

The findings of this study are related only to one service sector, namely retail banking services. It is quite possible that the results would have been different had a different service been incorporated. Testing the robustness of a scale to measure corporate reputation across different services leaves scope for future research.

The conclusion of this study is that the shortened 17-item instrument described here demonstrates superior evidence of construct validity compared to the original 31-item instrument proposed by Walsh and Beatty (2007). However, we believe that the original 31-item instrument, and particularly the operationalisation of the underlying constructs and the wording of the items used to measure these dimensions, are so flawed that there is an urgent need to start afresh in our efforts to find a valid instrument to measure a service firm's corporate reputation from a customer perspective.

References

- Alsop, R.J. 2004. *The 18 immutable laws of corporate reputation: Creating, protecting, and repairing your most valuable assets*. London: Kogan Page.
- Anderson, B. 2008. 'Into the tiger's liar', *Finweek*, 29 May:12-15.
- Bagozzi, R.P. 1980. *Causal models in marketing*. New York: John Wiley & Sons.
- Bagozzi, R.P. & Yi, Y. 1988. 'On the evaluation of structural equation models', *Journal of the Academy of Marketing Science*, **16**(1):74-94.
- Brown, T.J., Dacin, P.A., Pratt, M.G. & Whetten, D.A. 2006. 'Identity, intended image, construed image, and reputation: An interdisciplinary framework and suggested terminology', *Journal of the Academy of Marketing Science*, **34**(2):99-106.
- Bruce, P. 2008. 'The thick end of the wedge', *Business Day*, 12 May, p.10.
- Churchill, G.A., Jr. 1979. 'A paradigm for developing better measures of marketing constructs', *Journal of Marketing Research*, **16**(1):64-73.
- Collins, C. 2004. 'Move over, Enron. Wal-Mart is the new punching bag'. [Online] URL: <http://www.csmonitor.com/2004/0223/p01s02-usec.html>. Accessed July 2008.
- Darlin, D. 2006. 'Inside the news; H.P., Red-faced but still selling', *The New York Times electronic edition*. 1 October. [Online] URL: <http://query.nytimes.com/gst/fullpage.html?res=9401EEDDI1730F932A35753C1A9609C>. Accessed July 2008.

- Diamantopolous, A. 2005. 'The C-OAR-SE procedure for scale development in marketing: A comment', *International Journal of Research in Marketing*, **22**(1):1-9.
- Fombrun, C.J. Gardberg, N.A. & Sever, J.W. 2000. 'The reputation quotient: A multi-stakeholder measure of corporate reputation', *The Journal of Brand Management*, **7**(4):241-255.
- Gerbing, D.W. & Anderson J.C. 1996. 'An updated paradigm for scale development incorporating unidimensionality and its assessment', *Journal of Marketing Research*, **25**:186-92.
- Hartley, W. 2008. 'Public trust in institutions at a new low', *Business Day electronic edition*, 8 April. [Online] URL: <http://www.businessday.co.za/Articles/Content.aspx?id=46210>. Accessed July 2008.
- Hattie, J. 1985. 'Methodology review: Assessing the unidimensionality of tests and items', *Applied Psychological Measurement*, **9**(6):139-164.
- Hinkin, T. R. 1995. 'A review of scale development practices in the study of organizations', *Journal of Management*, **21**(5):967-988.
- Hoogland, J.J. & Boomsma, A. 1998. 'Robustness studies in covariance structure modeling: An overview and a meta-analysis', *Sociological Methods & Research*, **26**:329-367.
- Jöreskog, K.G. & Sörbom, D. 1988. LISREL. Scientific Software International, Inc. Lincolnwood, Chicago, Illinois, United States of America.
- Judd, C.M., Jessor, R. & Donovan, J.E. 1986. 'Structural equation models and personality research', *Journal of Personality*, **54**(1):149-198.
- Mackenzie, S.B. 2001. 'Opportunities for improving consumer research through latent variable structural equation modelling', *Journal of Consumer Research*, **28**(1):159-166.
- Malhotra, N.K, Peterson M. & Kleiser, S. 1999. 'Marketing research: A state-of-the art review and directions for the twenty-first century', *Journal of the Academy of Marketing Science*, **27** (Spring): 160-83.
- Mathews, C. 2008. 'PwC to respond to R&E claim over Kibbly cash', *Business Day, Companies and Markets section*, 24 April, p. 1.
- Murray, K.B. 1991. 'A test of services marketing theory: Consumer information acquisition activities', *Journal of Marketing*, **55** (January): 10-25.
- Murray, K.B. & Schlacter, J.L. 1990. 'The impact of services versus goods on consumers' assessment of perceived risk and variability', *Journal of the Academy of Marketing Science*, **18**(1): 51-65.
- Nunnally, J.C. & Bernstein, I.H. 1994. *Psychometric theory*. New York: McGraw-Hill.
- Peter, J.P. 1979. 'Reliability: A review of psychometric basics and recent marketing practices', *Journal of Marketing Research*, **16**(1):6-17.
- Peter, J.P. 1981. 'Construct validity: A review of basic issues and marketing practices', *Journal of Marketing Research*, **18**(2):133-145.
- Raykov, T. & Marcoulides, G. A. 2000. *A first course in structural equation modelling*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Rose, R. 2008. 'Cliffhanger'. *Financial Mail*, 30 May, pp. 23-36.
- Rossiter, J.R. 2002. 'The C-OAR-SE procedure for scale development in marketing', *International Journal of Research in Marketing*, **19**(4):1-31.
- Satorra, A. & Bentler, P.M. 1988. 'Scaling corrections for chi-square statistics in covariance structure analysis'. In *1988 Proceedings of the Business and Economic Statistics Conference of the American Statistical Association*. Alexandria, Virginia, pp.308-313.
- Satorra, A. & Bentler, P.M. 1994. 'Corrections to test statistics and standard errors in covariance structure analysis'. In Von Eye, A. & Clogg, C.C. (eds.). *Latent variable analysis: Applications to developmental research*. Newbury Park: Sage, pp.399-419.
- Steenkamp, J-B.E.M. & Trijp, H.C.M. 1991. 'The use of LISREL in validating marketing constructs', *International Journal of Research in Marketing*, **8**(4):283-299.
- Tull, D.S. & Hawkins, D.I. 1993. *Marketing research: Measurement and method*. 6th Edition. New York: Macmillan.
- Walsh, G. & Beatty, S.E. 2007. 'Customer-based corporate reputation of a service firm: Scale development and validation', *Journal of the Academy of Marketing Science*, **35**(1):127-143.

**APPENDIX A
FINAL 17-ITEMS**

CUSTOMER ORIENTATION

- | | |
|--------|--|
| CUST 1 | My bank has employees who are concerned about customer needs |
| CUST 2 | My bank has employees who treat customers courteously |
| CUST 3 | My bank is concerned about its customers |

EMPLOYEE RELATIONS

- | | |
|----------|---|
| EMPLOY 2 | My bank seems to treat its people well |
| EMPLOY 3 | My bank seems to have excellent leadership |
| EMPLOY 4 | My bank has management who seems to pay attention to the needs of its employees |
| EMPLOY 7 | My bank seems to be well-managed |

RELIABLE AND STRONG COMPANY

- | | |
|-------|--|
| REL 4 | My bank looks like a good investment |
| REL 5 | My bank seems to make financially sound decisions |
| REL 8 | My bank seems to have a clear vision of its future |

QUALITY PRODUCTS AND SERVICES

- | | |
|--------|---|
| QUAL 3 | My bank stands behind the services that it offers |
| QUAL 4 | My bank develops innovative services |
| QUAL 5 | My bank offers services that are good value for money |
| REL 2 | My bank seems to recognise and take advantage of market opportunities |

SOCIAL RESPONSIBILITY

- | | |
|----------|--|
| SOCIAL 1 | My bank seems to make an effort to create new jobs |
| SOCIAL 2 | My bank would reduce its profits to ensure a clean environment |
| SOCIAL 3 | My bank seems to be environmentally responsible |