



2007-02-01

# The Effect of Different Inputs to Factor Analysis: an Example using Service Quality in UK Branch Banking

Joseph Coughlan

*Dublin Institute of Technology*, joseph.coughlan@dit.ie

Estelle Shale

*University of Warwick*, estelle.shale@wbs.ac.uk

Robert Dyson

*University of Warwick*, robert.dyson@wbs.ac.uk

Follow this and additional works at: <http://arrow.dit.ie/buschmarcon>



Part of the [Management Sciences and Quantitative Methods Commons](#), and the [Marketing Commons](#)

## Recommended Citation

Coughlan, J., Shale, E., Dyson, R.: The Effect of Different Inputs to Factor Analysis: An Example using Service Quality in UK Branch Banking, American Marketing Association, San Diego, February, 2007.

This Conference Paper is brought to you for free and open access by the School of Marketing at ARROW@DIT. It has been accepted for inclusion in Conference papers by an authorized administrator of ARROW@DIT.

For more information, please contact [yvonne.desmond@dit.ie](mailto:yvonne.desmond@dit.ie), [arrow.admin@dit.ie](mailto:arrow.admin@dit.ie), [brian.widdis@dit.ie](mailto:brian.widdis@dit.ie).



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 3.0 License](#)



# **The Effect of Different Inputs to Factor Analysis: An Example using Service Quality in UK Branch Banking**

## **Abstract**

Factor analysis has long been used in service quality research to understand the dimensions of the construct. This research reinvestigates this construct using two different methodologies (classical test theory and item response theory) in order to assess the homogeneity of the dimensions across the methodologies in a retail branch banking sample taken from a larger network in the UK. The findings show that the two methodologies give different results. Furthermore the choice of correlation matrix to input into Confirmatory Factor Analysis may be more important than is currently thought in the literature as they give different results in this sample.

## **Introduction**

Many of the concepts that marketing, and indeed many social science researchers, use cannot be directly measured (Weisberg 1984) and thus are often described as being latent variables. These 'latent' variables are then measured using a set of observed or manifest variables using a survey methodology typically using Likert or semantic differential scales. Using the measurement typology derived by Stevens (1951), much of the data, particularly at customer level, are ordinal in nature. This implies that the appropriate statistics for use with such data are limited, since the assumptions of much of the standard techniques require at least interval level data. It must be acknowledged that not all social science measurement theorists agree on this issue, for a discussion see Gaito (1980; 1986).

The aim of this paper is to investigate and compare methods for dealing with ordinal data using a relatively simple model from the field of service quality by Grönroos (1984). Commensurate with the extant research in the field, the multivariate method used in this paper will be factor analysis. The first section of the paper reviews the different methods of ordinal factor analysis beginning with the classical test theory

(CTT) approach and continuing with an exploration of Item Response Theory (IRT) for factor analysis. The next section briefly considers the service quality literature in the light of the data being used in this paper. Following on from this, the data collection method is described. Analysis of the data using both the CTT and IRT approaches is then presented. A discussion of the results follows this section. Finally some conclusions, practical recommendations and directions for future research are offered.

### **Classical Test Theory (CTT) Approach**

The CTT approach to factor analysis as described by Bartholomew et al. (2002) uses correlation (or covariance) matrices in the development of a factor structure. The most common method is to use the Pearson product moment correlation coefficient (known as Pearson's  $r$ ). Unfortunately this requires that the data be at least interval in nature and as a result is not suitable for ordinal data. Theory testing in service quality beginning from the initial paper of Parasuraman et al. (1988) has relied on Pearson correlations as inputs to factor analysis. This research seeks to consider the use of other correlation matrices within a Confirmatory Factor Analysis (CFA) framework.

The use of different correlation matrices for ordinal data is common in psychological research but has not as yet been fully accepted in marketing research. Researchers such as Flora and Curran (2004) have used matrices other than the standard Pearson's  $r$ . Babakus et al. (1987) studied the use of alternative correlation measures as an input to Maximum Likelihood Confirmatory Factor Analysis. They took Pearson's  $r$ , the polychoric, Spearman's  $\rho$  and Kendall's  $\tau$ -b. They found that the polychoric performed worst with respect to convergence rates and improper solutions, especially

when the data was extremely skewed (skewness = 1.5), though this was mainly from cases with small samples and low factor loadings. However the polychoric performed best in the accuracy of parameter estimates and also estimated standard errors the most accurately. The polychoric correlation coefficient was also found to have led to rejection of correctly specified models much more often than the other three measures of correlation. In terms of goodness of fit, the polychoric was outperformed by the others but the differences were negligible other than in the presence of skewness. They suggest that “the researcher is well advised to analyse polychoric correlations when the data are ordinal” (Babakus et al. 1987: 227).

Rigdon and Ferguson (1991) note that when ordinal data are analysed by Maximum Likelihood Factor Analysis procedures the resulting estimates of parameters are biased and that the problem exists for sample sizes up to 500. They suggest the use of the polychoric correlation as a solution to this problem. The estimation of the polychoric correlation coefficient assumes that the unseen underlying variables are continuous and have a bivariate normal distribution and is estimated using maximum likelihood procedures (Olsson 1979). Rigdon and Ferguson (1991) suggest that reported better performance in using the polychoric correlation coefficient may be down to the effects of using maximum likelihood and other methods should be tested. They developed simulations to test the performance of the coefficient on different sample sizes, different fitting methods and different shapes of the distribution of the ordinal data. They found that an adequate sample size provided sufficient insurance against problems in estimation of parameter estimates. Fit of the models worsened as the distribution of the ordinal variable was more skewed. They acknowledge that the results are specific to their model and their simulated data.

Flora and Curran (2004) note that the simulation studies they reviewed showed that as the skewness and kurtosis of the observed ordinal variables increased, the estimation deteriorated even though statistical theory of CFA with polychoric correlations makes no explicit assumptions about the skewness and kurtosis of observed ordinal variables. They find in their study that the polychoric correlation coefficient is not robust to extreme violations of non normality.

Other correlation matrices are available that are suitable for ordinal data. These have been relatively under-researched and much attention has been given to the polychoric correlation over these. This research aims to redress this by using two other correlation coefficients to assess the factor structure and for comparison purposes: Spearman's rho and Kendall's tau-b following the work of Babakus et al. (1987).

Spearman's rho is a special case of Pearson's product moment correlation. Instead of analyzing all the data, the data is ranked and analysis of ranks is carried out. It accommodates many of the causes of distortion (e.g. outliers and non-normality) that cause problems for Pearson's r (Lane 2003). Kendall's tau is a measure of correlation between two ordinal variables. It also uses ranked data but it measures the number of pairs that are concordant. Concordance occurs when for two pairs of data  $(x_i, y_i)$  and  $(x_j, y_j)$ ,  $x_i > x_j$  and  $y_i > y_j$ . If  $x_i > x_j$  and  $y_j > y_i$ , then the pairs are said to be non concordant. Kendall's tau is different from Spearman's rho in that the tau value represents a probability that the data are in the same order versus they are not in the same order. Tau-b is the difference between concordant and non concordant pairs

divided by a term representing the geometric mean between the number of pairs not tied.

### **Item Response Theory (IRT)**

Direct methods of dealing with Likert scale data are uncommon in marketing but are an area of significant development in psychological measurement. The theoretical underpinning for these methods is Item Response Theory (IRT). Bartholomew et al. (2002) note that the application of these methods to factor analysis is at the edge of the research frontier. IRT is an approach to the analysis of ordinal data and is an extension of logit /probit models (Bartholomew et al. 2002). It differs from the other approach (the Classical Test Theory approach) in that no correlations are required between the variables and that it is a Full Information Maximum Likelihood (FIML) approach (Moustaki 2001). Bartholomew et al. (2002) note that use of ordinal data as if it were interval leads to biased estimates of factor loadings. They also make the point that FA based on correlation matrices other than the standard Pearson's  $r$  may be useful but that the method developed by use of IRT is better. Moustaki (2000) developed a general class of such models which she named the Proportional Odds Model (POM). Jöreskog and Moustaki (2001) further extended the POM approach.

In IRT probabilities are specified for each category from the data or they can be pre-specified from the researcher. Basically if there are  $m_i$  categories for variable  $i$  then the response probabilities (for  $m_i = 5$ ) are as per table 1:

<i>Category</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
<i>Response</i>	$\pi_{i(1)}(y)$	$\pi_{i(2)}(y)$	$\pi_{i(3)}(y)$	$\pi_{i(4)}(y)$	$\pi_{i(5)}(y)$

Table 1: Categories and Responses in IRT

As these are probabilities they all add to 1.0. Therefore  $\pi_{i(s)}(y)$  is the probability that given  $y$ , a response falls into category  $s$  for a given variable  $i$ . As IRT commonly uses the logit model (Moustaki 2003) and as the logit model is a binary model then these individual probabilities must be made into a binary format. These models (logit/probit) are known as link functions in that they are monotonically increasing functions that map  $(0, 1)$  onto  $(-\infty, \infty)$ . Bartholomew et al. (2002) then use the binary logit model for all possible divisions of the  $m_i$  categories into two groups based on the equations above. This logit model is written as follows:

$$\log \left[ \frac{1 - \gamma_{i(s)}(y)}{\gamma_{i(s)}(y)} \right] = \alpha_{i(s)} + \sum_{j=1}^q \beta_{ij} y_{ij}$$

The intercept  $\alpha$  for each category ( $s$ ) is calculated and they will be strictly ordinal. This model is also known as the Proportional Odds Model (POM) in that “in the one factor case, the difference between two cumulative logits [left hand side of last equation] for two persons with factor scores  $y_1$  and  $y_2$  is proportional to  $y_1 - y_2$ ” (Bartholomew et al. 2002: 211). Another issue is that the factor loadings ( $\beta$ ) remain the same across the categories, so the discriminating power of the model does not depend on the split between the two models (Jöreskog and Moustaki 2001). The basic assumptions as outlined by Bartholomew et al. (2002) are that the latent variables are independent and normally distributed with mean zero and variance one; and the responses to the ordinal items are independent conditional on the latent variables.

Goodness of fit characteristics can be calculated in terms of chi-squared statistics. The main problem with these methods is that they assume that all response patterns are in the data. However this is usually not the case. For a four variable model with

five categories this generates  $4^5$  (=1,024) unique categories. So therefore a sample size of 1,024 would be required just to have one response per category. Usually therefore many of the response categories have no responses. As a result the fit of the model will not be able to be properly assessed.

Jöreskog and Moustaki (2001) suggest the use of a rule of thumb to assess fit. They suggest that the values inside the chi-squared residual (S) for the two-way marginals of items should be less than  $(4 * m_i * m_j)$ , though Bartholomew et al. (2002) change this to three. However this still suffers from the problem of not all response categories having data. There are many advances in this field and the popularity of the method will definitely increase when this issue of fit is solved and when software is commercially available. Some steps have been made in this direction by Moustaki (2003) however this is not commercially available at this time. The method of Bartholomew et al. (2002) was chosen for this research as it built upon a strong foundation in logit analysis and furthermore because software was freely available that could work with datasets of the size encountered in this research and could estimate models of the size required.

### **Service Quality**

Service quality has been described as one of the most intriguing constructs in marketing theory (Roest and Pieters 1997). This is due to the elusive nature of the two concepts that it brings together. Service quality is seen as a key performance measure in its own right. It is seen as crucial to organizational success and survival for all organizations, but especially those with a high service component to their products.

Literature in the area of service quality and services marketing has alluded to two schools of thought: the Nordic (Scandinavian) school and the American school (Mels et al. 1997). These are best typified by a number of key authors. The Nordic school is concerned with the conceptualization of service quality and not as interested in its measurement. The basic tenets of the work lie in the distinction between different types of service quality. The major paper in this school which is typified by conceptualization rather than measurement work is that of Grönroos (1984). It deals with three aspects of quality:

- Technical Quality of the Outcome
- Functional Quality of the Service Encounter (Interaction)
- The Corporate Image

It accounts for the quality of the outcomes of the service process, tangible and intangible and also explicitly accounts for the interactive quality of the encounter. It also makes the important point that the image of the firm, an indeed of the brand, plays an important part in the assessment of quality. A recent comment by Grönroos (2001) casts some doubt on this whole area. He notes that the image concept in the original research was meant to incorporate a dynamic aspect since customers have continuous interactions with the firm. He also notes that the “technical and functional quality dimensions of a service replace the product features of physical product, nothing else” (Grönroos 2001: 151). He even goes so far as to say that “quality as such should not be measured” (Grönroos 2001: 151). He feels that the functional and technical issues highlighted in his original research are features of the service rather like packaging is a feature of the physical product.

The American school of thought, which has by far the most published research, is centered on the work of Parasuraman, Zeithmal and Berry. The conceptualization of quality here is somewhat different. The American school is built on the service-gaps model (Parasuraman et al. 1985). It posits that service quality is the result of a gap between what the customer perceives the providers provide and the customers expect.

It must be noted at this stage that the fundamental difference between the American and Nordic schools is that the American school has focused on the measurement of the process elements of the service and has neglected the outcome and more dynamic elements that the Nordic school sees as equally important. Although the work of the Nordic school is central to the development of a definition of service quality, it has been comparatively neglected in favor of the American school. The American school has the advantage that it actively measures service quality, while the Nordic school considers conceptualization issues only. It is interesting to consider at this point that perhaps the main problems with the measurement of service quality lie with its inadequate conceptualization and that a bringing together of the two approaches may be beneficial. This has been considered in the work of Mels et al. (1997) and Brady and Cronin (2001) among others. This research will return to the Nordic school of service quality and concentrate on developing this research stream.

### **Data Collection**

Data was collected by a major UK bank in a monthly telephone assisted survey of their customers. Customers were randomly sampled from the customer database of those that had given permission to the bank to contact them. Customers were asked seven questions related to service quality at their branch, as per table 2, using a five

point Likert scale labeled ‘Excellent’, ‘Very Good’, ‘Good’, ‘Fair’ and ‘Poor’, provided they had visited their branch in the last four weeks. This study uses data from a group of sixteen branches in the same city to consider the factor structure in the months of February to May 2002 from a larger sample.

<i>Question Code</i>	<i>Description<sup>1</sup></i>
sq1a	Queuing
sq1b	Cleanliness & Tidiness
sq2a	Understandability of Staff
sq2b	Politeness of staff
sq2c	Efficiency of staff
sq2d	Staff knowledge
sq2e	Staff treated customer as valued

Table 2 Service Quality Questions

Branch level data was too sparse to do the analysis and a group of branches from the same metropolitan area were taken. Data from four months were combined again due to a lack of data at group level on a monthly basis. Differences of means were explored to see if the data could be pooled in this way and with the exception of the first question on queuing, there were no significant differences across the months or the branches at the 1% level.

Given the structure of the questions, two factors are expected from the analysis. The first factor should be the first two variables which are clearly related to the technical

---

<sup>1</sup> Although the actual questions are available to the researcher, the above descriptions were chosen as confidentiality was ensured to the Bank as a condition of getting access to the data.

quality of the outcome while the remaining five cover the more interactive elements of the service (Grönroos 1984). Due to all the branches belonging to the same network, the corporate image aspect of Grönroos (1984) is controlled for in this piece of research.

## **Analysis**

This section of the paper focuses on the analysis of the data collected. The first subsection will concentrate on the univariate analysis and the exploratory factor analysis (EFA). The next section will consider the results from the CTT approaches and the final section is devoted to the IRT results.

## **Univariate and EFA**

Normality is traditionally assumed for this type of data as it is a prerequisite for many statistical tests. A one sample Kolmogorov-Smirnov Test for Normality was applied to each of the variables which showed the data was non-normal. The variables are generally positively skewed (however none above 1.5) and leptokurtic, with the exception of sq1a which is slightly platykurtic. This is to be expected given the generally positive impressions of the bank that their customers have. Standard textbooks would now suggest that the data be transformed to be more normal but this shall not be carried out here as the methods of ML are suitably robust to deal with departures from normality (Flora and Curran 2004; Rigdon and Ferguson 1991).

Exploratory FA was carried out in SPSS using ML and a one factor structure emerged explaining 55.24% of the variance however the chi-squared test was not significant. The communalities of sq1a and sq1b were very low (0.275 and 0.238 respectively)

and therefore these variables were removed. The factor analysis was re-done with the five variables and the chi-squared test was significant, all loadings were greater than 0.74 and it explained 67.04% of the variance. Given the literature suggests that these questions should break into two distinct factors, this is an interesting result.

A two factor solution was specified for the data also using ML again using SPSS. Although this result had a significant Chi-squared statistic, it gave rise to Heywood cases (where the communalities were higher than 1 in the process of estimation) and thus the results should be interpreted with caution. The next stage of the analysis was to consider using Confirmatory Factor Analysis (CFA).

#### **CFA using LISREL**

CFA was carried out using LISREL. As specified earlier in the literature section, a number of different correlation matrices were tested: the Pearson Product-Moment; Spearman; Kendall tau-b and the Polychoric. Although the individual factor loadings varied, they followed a similar pattern as can be seen from table 3. Factor loadings for the first factor were generally lower but still quite acceptable being uniformly above 0.6000. The highest loadings were found using the Polychoric which links to previous research by Babakus et al. (1987).

<b>Method</b>	<b>sq1a</b>	<b>sq1b</b>	<b>sq2a</b>	<b>sq2b</b>	<b>sq2c</b>	<b>sq2d</b>	<b>sq2e</b>
<b>Polychoric</b>	.7000	.7100	.9200	.8400	.9200	.8400	.8600
<b>Kendall</b>	.6000	.6400	.6860	.7500	.8400	.7600	.7700
<b>Pearson</b>	.6700	.6200	.8700	.7500	.8700	.7900	.8000
<b>Spearman</b>	.6400	.6600	.8700	.7700	.8500	.7900	.7900

Table 3: CFA estimates in LISREL

Unfortunately however the fit of the models was poor. Only the model using Kendall's tau-b had a significant Chi-squared value ( $p = 0.05385$ ). It also had an RMSEA of 0.045, though the P value for the test of close fit was 0.56. The family of Goodness of Fit Indices (i.e. Normed Fit Index, Comparative Fix Index) were all above 0.98 with the exception of the Parsimony Normed Fit Index which was 0.61. Finally the RMR was 0.022 which shows a good level of fit. The model using the Spearman rank correlation coefficient had an RMSEA of 0.06 which is within the bounds suggested by Ullman (1996), and the Pearson method likewise had an RMSEA of 0.061.

Upon inspection of the LISREL output using the polychoric correlation coefficient, the modification indices suggested that sq2b and sq2c should also load onto the first latent variable. Although this was tested, it did not significantly improve the RMSEA or the other fit indices. In the analysis of the LISREL output for the Pearson correlation coefficients, the link between sq2b and the first latent variable was also suggested. In this case, adding a link improved the model fit significantly with the RMSEA dropping to 0.041 and the p-value of the Chi-Squared reaching 0.04984.

However, the strength of the relationship was low (only 0.28). No Modification Indices other than allowing the errors to correlate appeared in the LISREL output relating to the model using the other two sets of correlation coefficients.

### Item Response Theory

A two factor model was fitted using GENLAT (Bartholomew et al. 2002) and the factor loadings were very different to those above. The first factor included all the items excluding sq1b which solely identified the second factor as can be seen from table 4.

Item	Fact1	Fact2
sq1a	<i>0.7052</i>	0.4230
sq1b	0.5311	<i>0.7802</i>
sq2a	<i>0.9401</i>	0.2676
sq2b	<i>0.8927</i>	0.3433
sq2c	<i>0.9533</i>	0.1902
sq2d	<i>0.9384</i>	0.1640
sq2e	<i>0.9570</i>	0.0663

Table 4 GENLAT Results

This is an interesting result and is different to the CTT approaches discussed above. However the problem with this method is testing the fit of the model. This is a seven variable model with five response categories so therefore the minimum sample size is 16,807 ( $7^5$ ). This paper has a sample size of 349 and analysis has shown that there are only 172 distinct response patterns of which the majority (67.74%) can be classified into five distinct response patterns.

The heuristic statistic for model fit (Joreskog and Moustaki 2001) is calculated for each of the pairs of categories for each of the variables. For this data set with five categories, the statistic should be less than 100 for each observation. Unfortunately this is violated twice in the result, though the level of violation is not high (just over 111). A one-factor version of the model failed to converge in GENLAT.

## **Discussion**

The results showed that using different methods of dealing with ordinal data in factor analysis gave different results, a finding that supports previously conducted research using artificial data and simulation (Babakus et al., 1987, Flora and Curran, 2004). This research extends these findings to the consideration of IRT as an alternative method to the traditional CTT approaches.

Using the CTT approach the CFA confirmed the pre-supposed results that the variables would split into two factors. However the CFA was only significant using Kendall's tau-b correlation coefficient using the Chi-Squared as a test statistic though results using the Spearman's rho were also significant when considering RMSEA as a guide (Ullman 1996). These are interesting results as when the two most frequently used correlation coefficients, the Pearson product-moment and the polychoric, were used, the CFA model was not supported. This has important implications for researchers dealing with Likert type data in marketing research.

The IRT approach gave a different result yet again and although a two factor solution was found, it did not support the Grönroos (1984) model. The outcomes factor was specified by only one item but the more obvious outcome item loaded onto the

interaction factor. This may be due to a problem with the fit of the model from the heuristics but the violation did not occur with this item in any case.

Comparison of the two approaches is useful and brings another dimension to the analysis of service quality data that is currently not present in the data. It questions the reliability of extant research and suggests that the choice of correlation coefficient and indeed theoretical frame of methodological analysis should be chosen with more care by future researchers.

### **Limitations**

Although the data set is limited in that it is the grouping of data from different branches over a number of time periods, it is useful in considering the factor structure of the data. Also the full battery of SERVQUAL-type questions was not asked of the respondents due to time constraints. Unfortunately the IRT data could not be adequately tested and although it did violate the heuristic test suggested by Jöreskog and Moustaki (2001), it still showed how some interesting results. Finally the lack of unified software to deal with all these issues meant that correlations had to be transferred from SPSS to LISREL with some consequent loss of accuracy.

### **Conclusions**

Notwithstanding the prevalence of Likert scale data in marketing, there is a remarkable lack of attention to the problem of how to deal with this data as ordinal. There has been a lot of debate on the issue (Gaito 1980; Hand 1996) and philosophically the issue is still unresolved. The lack of methods to deal with ordinal data has also contributed to this lack of attention. This research has moved to fill this

gap in the literature and to use extant methods of dealing with ordinal data from different literatures in a marketing context.

The outcomes in this regard are important for both practitioners and academia. It offers the practitioner a method of using ordinal data as ordinal using available software. For the academic, it also offers a methodology and shows that there are distinct differences between the methods chosen to do factor analysis on ordinal data. It further tests the analysis methodologies of Babakus et al. (1987) and Flora and Curran (2004) in that it applies their research to actual customer data rather than simulated data.

### **Recommendations for Future Research**

From a measurement point of view, the approaches followed in this paper for factor analysis were problematic in that the information from both the CTT and IRT approaches could not be fully compared. In fact the data that was available brought out interesting differences in the data set that could be further explored. Furthermore the lack of a test statistic, such as a Chi-squared, for sparse data sets as identified in the paper is an area for future investigation. For the IRT stream of research to gain acceptance in the marketing literature where it has much to contribute, a series of test statistics that work with the specific characteristics of the data being collected is required.

There may be branch or branch manager effects on the service quality variables that did not arise as a result of the way the analysis was carried out. A multi-level modelling approach (Hox 1995) could be used to consider if there are higher-level

effects. Future research will consider a wider sample of data and also seek to test the classic SERVQUAL instrument in the same manner as this paper.

## References

- Babakus, Emin, Carl E. Jr. Ferguson, and Karl G. Joreskog (1987), "The sensitivity of Confirmatory Maximum Likelihood Factor Analysis to violations of measurement scale and distributional assumptions," *Journal of Marketing Research*, 24 (May), 222-28.
- Bartholomew, David J., Fiona Steele, Irini Moustaki, and Jane I. Galbraith (2002), *The Analysis and Interpretation of Multivariate Data for Social Scientists*. Boca Raton, Florida: Chapman & Hall.
- Brady, Michael K. and J. Joseph Jr. Cronin (2001), "Some new thoughts on conceptualising perceived service quality: A hierarchical approach," *Journal of Marketing*, 65 (July), 34-49.
- Flora, David B. and Patrick J. Curran (2004), "An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis with Ordinal Data," *Psychological Methods*, 9 (4), 466-91.
- Gaito, John (1980), "Measurement scales and statistics: resurgence of an old misconception," *Psychological Bulletin*, 87 (3), 564-67.
- (1986), "Some issues in the measurement-statistics controversy," *Canadian Psychology*, 27 (1), 63-68.
- Grönroos, Christian (2001), "The perceived service quality concept - a mistake?" *Managing Service Quality*, 11 (3), 150-52.
- (1984), "A service quality model and its marketing implications," *European Journal of Marketing*, 18 (4), 36-44.

- Hand, D. J. (1996), "Statistics and the theory of measurement," *Journal of the Royal Statistical Society: Series A*, 159 (3), 445-92.
- Hox, J.J. (1995), *Applied Multilevel Analysis*. Amsterdam: TK-Publikaties.
- Jöreskog, Karl G. and Irini Moustaki (2001), "Factor Analysis of Ordinal Variables: A comparison of three approaches," *Multivariate Behavioral Research*, 36 (3), 347-87.
- Lane, David M. (2003), "HyperStat Online Textbook." Available: <http://davidmlane.com/hyperstat/>
- Mels, Gerhard, Christo Boshoff, and Deon Nel (1997), "The dimensions of service quality: the original European perspective revisited," *The Services Industries Journal*, 17 (1), 173-89.
- Moustaki, Irini (2003), "A general class of latent variable models for ordinal manifest variables with covariate effects on the manifest and latent variables," *British Journal of Mathematical and Statistical Psychology*, 56, 337-57.
- (2000), "A latent variable model for ordinal variables," *Applied Psychological Measurement*, 24 (3), 211-23.
- (2001), "A review of exploratory factor analysis for ordinal categorical data," in *Structural Equation Modeling: Present and Future*, Robert Cudeck and Stephen Du Toit and Dag Sorbom, Eds. Lincolnwood, Illinois: Scientific Software International.
- Olsson, Ulf (1979), "Maximum likelihood estimation of the polychoric correlation coefficient," *Psychometrika*, 44 (4), 443-60.
- Parasuraman, A., Leonard L. Berry, and Valerie A. Zeithmal (1985), "A conceptual model of service quality and its implications for future research," *Journal of Marketing*, 49 (Fall), 41-50.
- (1988), "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality," *Journal of Retailing*, 64 (Spring), 12-40.

Rigdon, Edward E. and Carl E. Jr. Ferguson (1991), "The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data," *Journal of Marketing Research*, 28 (Nov), 491-97.

Roest, Henk and Rik Pieters (1997), "The nomological net of perceived service quality," *International Journal of Services Industry Management*, 8 (4), 336-51.

Stevens, S. S. (1951), "Mathematics, Measurement and Psychophysics," in *Handbook of Experimental Psychology Part I*, S. S. Stevens, Ed. New York: Wiley.

Ullman, Jodie B. (1996), "Structural Equation Modelling," in *Using Multivariate Statistics*, Barbara G. Tabachnick and Linda S. Fidell, Eds. Boston: Allyn and Bacon.

Weisberg, Herbert F. (1984), "The Fundamentals of Data Analysis," in *Theory-Building and Data Analysis in the Social Sciences*, Herbert B. Asher and Herbert F. Weisberg and John H. Kessel and W. Phillips Shively, Eds. Nashville: University of Tennessee Press.