Application of the European Customer Satisfaction Index to Postal Services. Structural Equation Models versus Partial Least Squares

Christina O'Loughlin*, Germà Coenders†

Departament d'Economia,
Universitat de Girona

Girona, September 2002

Abstract

Customer satisfaction and retention are key issues for organizations in today's competitive market place. As such, much research and revenue has been invested in developing accurate ways of assessing consumer satisfaction at both the macro (national) and micro (organizational) level, facilitating comparisons in performance both within and between industries. Since the instigation of the national customer satisfaction indices (CSI), partial least squares (PLS) has been used to estimate the CSI models in preference to structural equation models (SEM) because they do not rely on strict assumptions about the data. However, this choice was based upon some misconceptions about the use of SEM's and does not take into consideration more recent advances in SEM, including estimation methods that are robust to non-normality and missing data.

In this paper, both SEM and PLS approaches were compared by evaluating perceptions of the Isle of Man Post Office Products and Customer service using a CSI format. The new robust SEM procedures were found to be advantageous over PLS. Product quality was found to be the only driver of customer satisfaction, while image and satisfaction were the only predictors of loyalty, thus arguing for the specificity of postal services.

Keywords: European Customer Satisfaction Index (ECSI), Structural Equation Models, Robust Statistics, Missing Data, Maximum Likelihood.

JEL classification: C13, C39, C51, H42, L89, M11.

* Heart Failure Unit, St. Michael's Hospital, DunLaoghaire Co. Dublin, Ireland. E-mail: drologuhlin@ hotmail.com

† Address: Departament d'Economia. Universitat de Girona, Campus de Montilivi, 17071 Girona, Spain. E-mail: germa.coenders@udg.es

1. Introduction

1.1. Origin and Uses of the Customer Satisfaction Index

Customer satisfaction has become a vital concern for companies and organizations in their efforts to improve product and service quality, and maintain customer loyalty within a highly competitive market place. In the last decade, a number of national indicators reflecting consumer satisfaction across a wide range of organizations have been developed (e.g., Sweden, Fornell, 1992; USA, Fornell et al., 1996; Norway, Andreassen & Lindestad, 1998a, 1998b; Denmark, Martensen, Grønholdt & Kristensen, 2000; European Union, ECSI Technical Committee, 1998). At the national level, the customer satisfaction index (CSI) is a nationwide gauge of how adequately companies, and industries in general satisfy their customers. In addition, CSI's can be used at the lower industry or even company level facilitating comparison of companies within an industry. These indicators complement traditional measures of economic performance (e.g., return on investment, profits and market shares) providing useful diagnostics about organizations, and their customers evaluations of the quality of products and services.

1.2. Factors within the ECSI/ACSI Model

The basic structure of the CSI model has been developed over a number of years and is based upon well established theories and approaches to consumer behaviour, customer satisfaction and product and service quality (see Fornell, 1992; Fornell et al., 1996). The structure of the CSI is continually undergoing review and subject to modifications. Although the core of the model is in most respects standard, there are some variations between the SCSB (Swedish), the ACSI (American), the ECSI (European), the NCSB (Norwegian) and other indices. For example, the image factor is not employed in the ACSI model although plans are underway to include this factor into this model (Johnson et al., 2001).

The CSI model consists of a number of latent factors, each of which is operationalised by multiple indicators. Customer satisfaction (SATI) can be defined as an overall evaluation of a firm's post-purchase performance or utilization of a service (Fornell, 1992). It is at the core of the CSI framework and is encased within a system of cause and effect running from the antecedents of overall customer satisfaction - expectations, image, perceived quality and value – to the consequences of overall customer satisfaction – customer loyalty and customer complaints. The obvious strength of this approach is that it moves beyond the immediate consumption experience and facilitates the study of the causes and consequences of consumer satisfaction. In fact, the primary objective of this structural approach is to explain customer loyalty.

Antecedents of customer satisfaction:

1) *Perceived Quality*: In 1996, the ACSI model was expanded to delineate two general types of perceived quality, product quality (hardware) and service quality (software/humanware) (Fornell et al., 1996). Perceived product quality (QUAL1) is the evaluation of recent consumption experience of products. Perceived service quality (QUAL2) is the evaluation of recent consumption experience of associated services like customer service, conditions of product display, range of services and products etc. This distinction between service quality and product quality is a standard feature of the ECSI model (Eklöf, 2000). Kristensen et al., (1999) demonstrated the importance of delineating these two aspects of perceived quality in a post office context. Both QUAL1 and QUAL2 are expected to have a direct and positive effect on overall customer satisfaction.
2) *Value (VALU)*: The literature in this area has recognised that customer satisfaction is dependent on value (Howard & Sheth, 1969). Value is the perceived level of product quality relative to the price paid or the "value for money" aspect of the customer experience. Value is defined as the ratio of perceived quality relative to price (Anderson et al., 1994). Value is expected to have a direct impact on satisfaction (Anderson & Sullivan, 1993; Fornell, 1992) and to be positively affected by perceived quality (both QUAL1 and QUAL2). To ensure that the effects of a price-quality relationship are not confounded, quality and value are measured relative to each other (Anderson et al., 1994).
3) *Image (IMAG)*: Image refers to the brand name and the kind of associations customers get from the product/brand/company. This construct was first introduced in the Norwegian Customer Satisfaction Barometer (NCSB) model (Andreassen & Lindestad, 1998a; Andreassen & Lindestad, 1998b). New research indicates that it is an important component of the customer satisfaction model (e.g., Martensen et al., 2000). It is expected that image will have a positive effect on customer satisfaction and loyalty. In addition, image has been modelled to have a direct effect on value (e.g., Kristensen et al., 1999; Martensen et al., 2000). The impact of quality on image (or vice versa) is not usually estimated. According to Johnson et al., (2001), image has been modelled to affect perceptions of quality (Andreassen & Lindestad, 1998a). However, in most research papers this affect is not modelled, thus we consider image and product and service quality to be all exogenous factors.
4) *Expectations*: Expectations refer to the level of quality that customers expect to receive and are the result of prior consumption experience with a firm's products or services. Johnson et al., (2001) noted that the effect of expectations is non significant in a number of industry sectors. Similarly, Martensen et al., (2000) showed that customer expectations of post office products and services in Denmark have a negligible impact on consumer satisfaction. Thus, the expectations construct was not included in this paper.

Consequences of consumer satisfaction:

1) *Complaints (COMP):* This factor refers to the intensity of complaints and the manner in which the company manages these complaints. It is expected that an

increase in customer satisfaction should decrease the incidence of complaints (American Society for Quality, 1998; Fornell et al. 1996).

2) *Loyalty (LOYA):* Customer loyalty is the ultimate dependent variable in the model and is seen to be a proxy measure for profitability (Reichheld & Sasser, 1990). Increasing customer loyalty secures future revenues and minimises the possibility of defection if quality decreases. In addition, word-of-mouth from satisfied loyal customers embellishes the firm's overall reputation and reduces the cost of attracting new customers (Anderson & Fornell, 2000). Loyalty is measured by repurchase intention, price tolerance and intention to recommend products or services to others. It is expected that better image and higher customer satisfaction should increase customer loyalty. In addition it is expected that there is a reciprocal relationship between complaints and loyalty. When the relationship between customer complaints and customer loyalty is positive it implies that the firm is successful in turning customers who complain into loyal customers. Conversely, it is expected that when the relationship is negative the firm has not handled complaints adequately.

Our model is thus displayed in Figure 1.

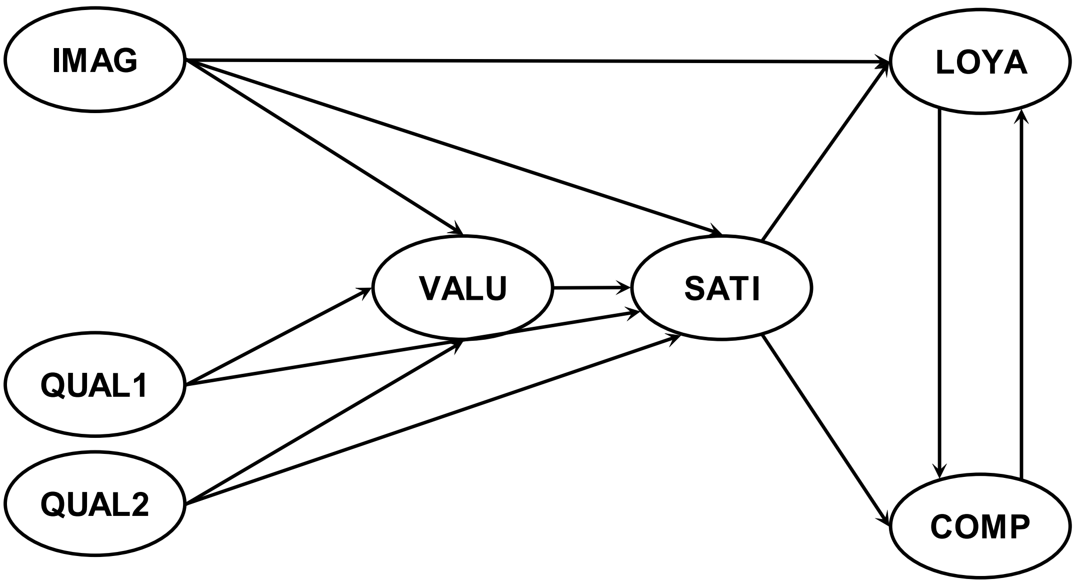

Figure 1. CSI model

1.3. Estimation Procedures

From the beginning, Partial Least Squares (PLS, Wold, 1975) has been suggested as a method for estimating the latent variable CSI models (Fornell, 1992). This recommendation was grounded on the argument that the other widely employed framework used to estimate relationships among latent variables (covariance structure analysis models, also called structural equation models –SEM– and sometimes LISREL models, after the name of the first commercial software that became available, developped by Jöreskog, 1973. See Bollen, 1989, Raykov and Marcoulides, 2000; or Batista-Foguet & Coenders, 2000 as introductory manuals) makes more strict assumptions on the data, mainly regarding normality. This recommendation can be traced back to previous PLS literature (e.g. Fornell & Cha, 1994) and seems to have been widely followed. The examples of SEM models applied to the CSI are few, they do not take advantage of the latest developments in SEM (Hackl, Scharitzer & Zuba, 2000) and often conclude with a recommendation to use PLS instead (Kristensen, Martensen, & Grønholdt, 1999).

PLS estimates have been reported to be biased (Cassel, Hackl & Westlund, 2000). In fact, from a theoretical point of view, this bias would only disappear under perfect reliability of the observed variables or under an infinite number of indicators per latent variable. However, the arguments that can be found from the advocates of PLS are various:

1) PLS is best suited as a prediction technique. However biased their estimates may be, they yield optimal predictions of the dependent variables from the observed explanatory variables. This property is shared with ordinary least squared regression, that is also biased under the presence of measurement error in the explanatory variables, but also yields optimal predictions even in this case.
2) SEM assumes that observations are independent and follow a multivariate normal distribution. PLS uses non-parametric inference methods (such as jackknifing) and is free of these assumptions.
3) SEM scores for the individuals on the latent variables are indeterminate.

Some of these arguments are fallacious in the CSI context and others have been rendered outdated as new developments have come into being in the SEM field.

1) More often than not, PLS is used in the CSI context with the aims of validating the measurement instruments in the CSI scales and of estimating and testing the effects on some dependent variables, purely predictive applications being rare in the literature. In order to perform a construct validation, unbiased estimates of factor loadings, factor correlations and regression slopes are needed. Factor correlations and regression slopes are reported to have a negative bias (Dijkstra, 1983), thus failing to reveal relationships among factors that are relevant for nomological validation. Factor loadings are reported to be overerestimated and tend to be similar even when true loadings are dissimilar (Widaman, 1993) thus obscuring convergent validation. Even Jacknife confidence intervals are of little use when they are built around a biased point estimate.
2) The issue of robustness of SEM estimates was sometimes misunderstood in early applications. Normality and independence of observations are *not* required for unbiasedness of maximum likelihood (ML) estimates, but are necessary for their efficiency and for providing correct standard errors and test statistics. However, even the later property is unaffected by non-normality if certain conditions are fulfilled (Satorra, 1990). Robust procedures in SEM were considered from an early stage. Browne (1984) suggested an alternative estimation method that is efficient and leads to correct test statistics under arbitrary distributions as long as sample sizes are large and

the number of variables included in the model is small to moderate (Muthén & Kaplan, 1989). Satorra (1990, 1992, 1993) and Satorra and Bentler (1994) developed robust standard errors and test statistics under arbitrary distributions when using the still consistent standard ML estimation method, a solution found to be advantageous for smaller sample sizes or larger models (Fouladi, 2000). Muthén and Satorra (1995) extended the former robust statistics to complex samples, the commonest source of dependence among observations in survey data. Finally, statistical inference based on non parametric resampling methods (e.g. bootstrap) is equally possible for SEM as it is for PLS (see Nevitt & Hancock, 2001 for an evaluation of its performance).

3) PLS latent variable scores are computed in such a way that the reliability estimates of the indicators and the R^2 of the latent variable regressions are maximised (Fornell & Cha, 1994). This criterion is indeed objective but it can be considered to be responsible for the bias in PLS parameter estimates, as some of the parameters (error variances) are minimized as a part of the criterion function. In SEM, the parameters are first consistently estimated and next latent variable scores are estimated conditional on these parameters. The fact that several criteria or objective functions are available for doing this, and that they differ from the equally arbitrary PLS criterion, does not provide adequate reason to refer to these scores as indeterminate. The most common approach is regression, which is equivalent to maximising the reliability of the latent variable scores and is very appealing from a practical point of view whenever the aim of the research is the development of indices. As an alternative, latent variable scores can be computed so that their means, covariances and variances are as close as possible to the model estimated means, variances and covariances of the latent variables. Other methods are also available (see Bollen, 1989, p 305, and references therein).

Besides, SEM have also a number of advantages over PLS that must not be ignored:

1) SEM offer the possibility of testing the significance of omitted parameters (e.g. Bollen & Long, 1993), such as error covariances and loadings on more than one latent variable. Such parameters compromise validity if significant, as they show indicators to contain some common variance unrelated to the latent variable they are supposed to measure (Batista-Foguet & Coenders, 1998).
2) SEM offer more flexibility in the specification of the model parameters, like for instance the error covariances and loadings on more than one latent variable mentioned above (e.g. Coenders & Saris, 2000).
3) SEM make it possible to simultaneously estimate the same model in several populations.
4) SEM make it possible to use the statistical theory that is coupled with maximum likelihood estimation.
5) SEM allow researchers to constrain parameters to be equal to a given value or to a given linear or non-linear function of other parameters.

Finally, additional developments in SEM have rendered this approach far more general and flexible than PLS. These include:

1) Inclusion of observed variables with an ordinal measurement level (Muthén, 1984; Jöreskog, 1990; Coenders & Saris, 1995; Coenders, Satorra & Saris, 1997). PLS assumes interval measured observed variables.
2) Treatment of hierarchical two-level data structures, thus bringing together SEM and some forms of multilevel modelling (Muthén, 1994).
3) Inclusion of both latent and observed categorical variables, thus bringing together SEM, mixture models and latent class models (Muthén, 2001). PLS assumes interval measured observed and latent variables.
4) ML estimation with missing data (Allison, 1987; Muthén, Kaplan, & Hollis, 1987; Wothke, 2000; Muthén & Muthén, 2001; Graham, Taylor & Cumsille, 2001). Variants of the method that are robust to non normality have recently become available (Yuan & Bentler, 2000). PLS requires a complete data set, which makes only imputation, mean substitution and listwise deletion appropriate.

All these developments have reached the stage of wide acceptance and availability and are all currently available to the user in at least one of the last generation commercial SEM software programs, that are all very user friendly and no longer require matrix algebra knowledge from the user.

1.4. Plan of the Paper

This paper illustrates the use of both PLS and recent SEM estimation procedures that are robust to non-normality and consistent under incomplete data missing at random. The illustration utilises data from an administration of the ECSI questionnaire in an industry specific case, the postal services of the Isle of Man.

The paper will be organised as follows. First the data and questionnaire will be presented. Then seven alternative SEM estimation procedures will be discussed. Next a confirmatory factor analysis will be used to validate the questionnaire and will be estimated using the seven procedures and the results compared. The causal relationships between dimensions in the ECSI model will then be estimated using the best of the methods. After removal of non significant effects the results will be interpreted and the implications, both statistical and theoretical, discussed.

2. Method

2.1. The Measurement Instrument

A number of CSI questionnaire formats, including the ECSI and ACSI were examined to aid in the development of a survey instrument relevant to the Isle of Man Post Office.

In addition, Post Denmark supplied a copy of their questionnaire, the results of which were documented by Kristensen, Martensen and Grønholdt (1999). Excellence Ireland supplied copies of questionnaires used to assess CSI in the mobile phone and banking sectors in Ireland.

Latent variable	Observed variables
QUALITY OF PRODUCTS: HARDWARE (QUAL1)	Q1. Please rate the overall quality of products and services *Q2. Please rate the quality of products and services compared to the quality offered by similar companies* Q3. Does the overall quality of products meet your quality requirements? Q8. Could the quality of the products and services be improved? *Q12. In the last 6 months how often have things actually gone wrong with the products/services you use?*
QUALITY OF CUSTOMER SERVICE: HUMANWARE (QUAL2)	Q4. Please rate the overall quality of customer service *Q5. Please rate the quality of customer service compared to the quality offered by similar companies* Q6. Does the overall quality of customer service meet your quality requirements?
PERCEIVED VALUE (VALU)	Q7. Please rate the *quality* of products/services given the prices you pay Q10. Please rate the *prices* of products and services given the quality
IMAGE (IMAG)	*Q20. The PO is a financially sound company* Q21. The PO is a reliable and trustworthy company Q22. The PO is a customer-oriented company Q23. The PO provides a valuable service to the community Q24. The PO is innovative and forward looking *Q25. The PO uses modern technological equipment and procedures* *Q26. The PO premises are modern and visually appealing* *Q27. The PO is a competitive courier service provider* *Q28. The PO offers value for money*
CUSTOMER SATISFACTION (SATI)	Q15. Overall how satisfied are you with the PO products and services? Q16. How close is the PO to your *ideal* postal service provider? Q17. Considering your expectations, to what extent has the PO fallen short of or exceeded your expectations?
CUSTOMER LOYALTY (LOYA)	*Q11. If there was a competitive postal service provider that could offer the same range and quality of products/services as the PO how much would they have to reduce their prices for you to change provider?* Q18. If you required new postal/courier products or services how likely is it that you would choose the PO to provide you with these services? Q19. If asked for your advice, how likely is it that you would recommend the PO services you use to others?
CUSTOMER COMPLAINTS	*Q13. How many times have you complained (either formally or informally) to PO personnel?* *Q14. How well or poorly was your most recent complaint handled?*

Table 1: Isle of Man Post Office postal survey containing items pertaining to this article only. Questions in italics are absent from the final model

The first draft of the survey was issued to 30 individuals to examine the question wording and relevance. Feedback from this pilot study indicated that some questions were ambiguous, difficult to understand, or irrelevant in a Postal context and were reworded. The final questionnaire contained 28 questions pertaining to the global customer satisfaction index, 38 specific questions about products and services, and 3 demographic questions. Only questions relating to the global customer satisfaction index are discussed in this paper and are displayed in Table 1. Questions are rated on 1 to 10-point scale with the exception of Q11 and Q13.

2.2. Sample and Data Collection

The questionnaire was administered to an Isle of Man residential sample. Data were collected during the summer of 2001 using a postal data collection mode (for an overview of competing data collection modes and their cost/quality balance see for instance Groves, 1989). It was unnecessary to use screening questions as all residents on the Island are in receipt of a postal service. One thousand residential clients were randomly selected from The Isle of Man Electoral Register where all constituency areas were proportionally represented. As specified in ECSI/ACSI methodologies, responses were made on a 10 point scale. A total of 28% of this sample completed and returned the questionnaires, which is quite a good result for the postal administration method. On average, all constituencies were proportionally represented by this sample. Within each constituency, simple random sampling was used, which ensures approximate independence of observations.

2.3 Exploratory Analyses and Estimation Procedures

Univariate and bivariate sample statistics were first computed. The complaint dimension (Q13, Q14) was dropped from the model as 86% of respondents had not complained during the previous 6 months. Q12 was removed because of low correlations with other items in the QUAL1 dimension.

In order to prevent the effects of outliers on the results of multivariate statistical models (Barnett & Lewis, 1994), we computed the Mahalanobis distance to the mean vector (Mahalanobis, 1936) and removed 4 observations with extreme values.

As regards normality, many variables showed evidence of moderate to severe non-normality, as shown in Table 2.

Prior to performing estimation, variables (Q2, Q5, Q11, Q20, Q25, Q27) and cases with more than 25% missing data were deleted. Variables with larger proportion of missing data can be assumed to be meaningless to a portion of the population. Cases with a larger proportion of missing data can be assumed to correspond either to careless respondents or to infrequent users of the service. The effective sample size was 258. The distribution of the remaining missing data was as shown in Table 3, and amount to an overall rate of missingness of 4.2%.

Variable	Mean	St. Dev.	Skewness	Kurtosis
Q1	8.292	1.642	-0.965**	1.093**
Q3	8.396	1.700	-1.046**	0.691
Q4	8.488	1.682	-1.145**	1.143**
Q6	8.432	1.697	-1.124**	1.101**
Q7	7.663	1.832	-0.632**	0.096
Q8	7.184	2.576	-0.739**	-0.356
Q10	7.643	1.869	-0.508**	-0.305
Q15	8.436	1.671	-1.170**	1.122**
Q16	8.190	1.764	-0.871**	0.169
Q17	7.431	1.872	-0.618**	0.314
Q18	8.096	2.094	-1.618**	2.846**
Q19	8.213	1.895	-1.345**	1.877**
Q21	8.488	1.777	-1.478**	2.260**
Q22	7.833	2.109	-0.988**	0.673
Q23	9.044	1.398	-2.267**	7.212**
Q24	7.550	2.157	-0.768**	0.070
Q26	6.473	2.407	-0.333*	-0.585**
Q28	7.907	1.979	-1.156**	1.472**

Table 2: Univariate summary statistics and normality tests

** Skewness or kurtosis significant at the 1% level

* Skewness or kurtosis significant at the 5% level

Q1	Q3	Q4	Q6	Q7	Q8	Q10	Q15	Q16
3	4	2	3	10	29	15	1	11
Q17	**Q18**	**Q19**	**Q21**	**Q22**	**Q23**	**Q24**	**Q26**	**Q28**
27	18	14	3	8	2	28	10	6

Table 3: Number of missing values per variable

Missing data are treated in several alternative ways within the context of SEM.

1) *Listwise or pairwise deletion.* These procedures are only unbiased if the data are missing *completely* at random (see Little and Rubin, 1987). Data are said to be missing completely at random when the probability that a datum is missing is independent of any characteristic of the individual. Even under this unrealistic assumption, the first of the mentioned methods is highly inefficient and the second leads to biased standard errors (Enders, 2001; Enders & Bandalos, 2001). This method is feasible, though of course not recommended, both for SEM and for PLS.
2) *Mean substitution.* This method is known to bias variances and covariances even when data are missing completely at random (Graham, Hofer & Piccinin, 1994; Graham, Hofer & MacKinnon, 1996, Enders, 2001) and thus should not be used, though numerically speaking is feasible both for SEM and for PLS.
3) *Imputation.* This is a family of methods including regression imputation, hot deck imputation or EM imputation, in both their simple and multiple variants (Little and Rubin, 1987). This approach has the advantage of providing a complete data set on which standard estimation procedures could in principle be used. However, some of these imputation procedures (simple hot deck imputation and simple regression imputation) lead to biased estimates of factor correlations and residual variances.

Multiple imputation (Rubin, 1987) does not have these drawbacks but it is cumbersome to perform unless special software is available. Imputation can be justified both if the data are missing at random or completely at random. Data are said to be missing at random when the probability that a datum is missing depends only of characteristics of the individual that are observed (not missing). This method is feasible both for SEM and for PLS.

4) *Direct ML assuming that the data are normally distributed and missing at random.* This procedure is currently available in most of the latest commercial software packages for SEM like Mx (Neale et al., 1999), EQS 6.0 (Bentler, 2000), AMOS 4.0 (Aburckle & Wothke, 1999), LISREL 8.51 (Jöreskog et al. 2000; du Toit & du Toit, 2001) and MPLUS 2.1 (Muthén & Muthén, 2001). This procedure uses all available data to build a case per case likelihood function. It is consistent, efficient and leads to correct standard errors and test statistics if the data are normal and missing at random (Aburckle, 1996; Enders, 2001; Enders & Bandalos, 2001; Wothke, 2000). Yuan and Bentler (2000) describe other related approaches: a two-stage estimator based on the combination of the EM algorithm and a standard ML SEM estimator, and a minimum χ^2 estimator. Both approaches seem to have similar properties to the direct ML estimator. Except for the two-stage EM approach, these methods are only feasible for SEM.
5) A variant of the direct ML estimator with missing data described by Yuan and Bentler (2000) is robust to non-normality but assumes data to be missing completely at random and can be biased otherwise. This method is available at least in MPLUS 2.1 and EQS 6.0. The few studies conducted to date report that this method performs quite well, regarding both unbiasedness and standard errors, even when data are just missing at random (Enders, 2001). This method is only feasible for SEM.

When data are missing not at random (what is also called non-ignorable missing data) none of the procedures are consistent. This is the case when the probability that a datum is missing depends on characteristics of the individual that are missing, for instance on the same variable that is missing for the individual. However, ML is reported to be less biased than the alternative approaches (Muthén et al., 1987).

In this article we use SEM under six different approaches and, for comparative purposes, PLS is also applied using one of these approaches:

1) Direct robust ML assuming that the data are missing at random. This is the method that is correct under the weakest assumptions and is considered as a baseline for the comparison.
2) Direct standard ML assuming that the data are normally distributed and missing at random.
3) Robust ML on a complete data set obtained by a form of imputation that is related to simple hot deck imputation. For each case (recipient) with a missing datum, complete cases on a given set of related variables are selected. The case with the lowest distance to the recipient on the set of related variables is chosen as a donor for the missing datum. If more than one potential donor has the same minimum distance to the recipient, then the mean value for all donors is imputed but this is

only done if the variance of these values is lower than the variance of the imputed variable. This method is currently available in the PRELIS program from version 2 on (Jöreskog & Sörbom, 1993). The variables in the model with less than 5 missing cases were used as the set of related variables, together with external variables related to the frequency of use of a set of postal services, for which there were no missing data. 8 cases could not be imputed because no cases with complete data for the set of related variables were available or because the variance of the values of the donors was larger than the variance of the recipient variable. The final usable sample size was 250.

4) Standard ML on the above imputed data set.
5) Robust ML on a complete data set obtained by mean substitution.
6) Standard ML estimation on the mean substituted data set. This method is by far the worst of the six described so far and would be close to what was available for SEM in the early 1980's, that is when PLS were developed and recommended.
7) PLS on the mean substituted data set, with 100 jackknife resamples, using the path weighting scheme for the inner part of the model, selecting an outward measurement model for the outer part of the model and using the non-standardized data. PLS could have been used on the imputed data set as well, but the only aim of including it here is to compare the results of SEM and PLS on the same data set, for which purpose the mean substituted data are enough. Besides, the standard guidelines for the estimation of the ESCI model include the use of mean substitution.

The PLS estimations were carried out using the PLS-PC 1.8 program. All SEM estimations were carried out with the M-PLUS 2.1 program (Muthén & Muthén, 2001). Point estimates are the same for robust and non-robust SEM estimation methods. This is due to the fact that standard ML estimates are consistent under non-normality (if a proper missing value treatment is employed, of course) and will be used to save space when presenting the results. As regards robust test statistics for complete imputed data M-PLUS computes Satorra and Bentler's (1994) mean-and-variance adjusted χ^2 statistic and also performs an adjustment of the associated degrees of freedom. When the same is done for incomplete data, MPLUS computes Yuan and Bentlers T_2^* χ^2 mean scaled statistic (Yuan and Bentler, 2000), which does not involve any adjustment of the degrees of freedom. Both robust test statistics are robust to non-normality. Standard errors robust to non-normality are also provided. In the missing data case these robust standard errors are based on a sandwich procedure (Arminger & Sobel, 1990). Likelihood ratio χ^2 difference tests between nested models cannot be carried out with mean-and-variance adjusted χ^2 statistics but they can be obtained for mean scaled χ^2 statistics if some simple calculations are performed by hand (Satorra & Bentler, 1999). Let $T_2{}^*{}_0$ and $T_2{}^*{}_1$ be the mean scaled χ^2 statistics, c_0 and c_1 the scaling constants, and d_0 and d_1 the degrees of freedom for two nested models, of which Model 0 is more restrictive. This implies that the standard ML χ^2 statistics are $T_0= T_2{}^*{}_0c_0$ and $T_1= T_2{}^*{}_1c_1$. The robust χ^2 difference can be computed as:

$$Robust\ \chi^2\ difference = \frac{T_0\ -\ T_1}{\dfrac{d_0 c_0 - d_1 c_1}{(d_0 - d_1)}}$$

3. Results

3.1. Confirmatory Factor Analysis Model

As often recommended (e.g. Bollen, 1989, Batista-Foguet & Coenders, 2000), a confirmatory factor analysis (CFA) model was first fit to assess the quality of the indicators of the underlying factors and spot invalid items that do not only measure the intended factor but have other systematic sources. During this process, Q26 was dropped as it was the indicator of the image factor with the largest error variance as well as being involved in the three highest residual covariances, which suggests that the item measures other things than only image. Q28 was also dropped as it loaded on both the value and image factors, which was not unexpected given its wording. These problems can be detected in a straightforward manner in SEM, as the significance of omitted parameters is being tested. Had PLS been used, they would have gone unnoticed. The final CFA model is displayed in Figure 2.

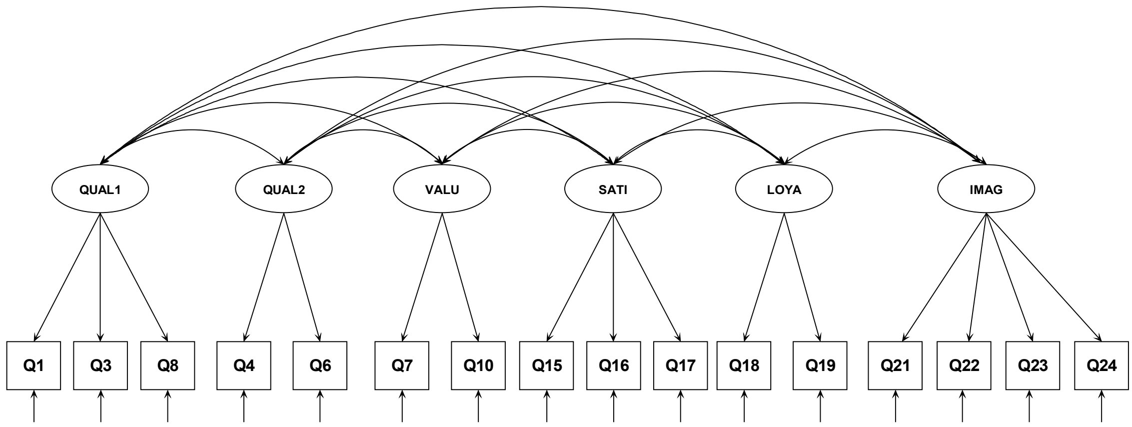

Figure 2: CFA model

Hot deck imputation

	Q1	Q3	Q4	Q6	Q7	Q8	Q10	Q15	Q16	Q17	Q18	Q19	Q21	Q22	Q23	Q24
Q1	2.70															
Q3	2.11	2.89														
Q4	1.80	1.93	2.83													
Q6	1.98	2.18	2.48	2.88												
Q7	1.87	2.18	1.82	2.09	3.36											
Q8	2.54	2.86	2.34	2.71	2.93	6.63										
Q10	1.83	2.16	1.79	2.06	2.67	2.97	3.49									
Q15	1.96	2.34	1.96	2.13	2.20	2.96	2.13	2.79								
Q16	1.79	1.95	1.67	1.80	1.81	2.68	1.95	2.17	3.11							
Q17	1.76	2.05	1.68	1.94	2.09	2.87	2.23	2.24	1.94	3.50						
Q18	1.64	1.94	1.57	1.84	1.87	2.37	1.67	1.96	1.97	1.95	4.38					
Q19	1.90	2.18	1.87	2.09	2.06	2.63	2.16	2.19	2.02	2.13	2.71	3.59				
Q21	1.72	1.98	1.74	1.86	1.89	2.51	2.00	1.93	1.82	1.92	1.79	2.17	3.16			
Q22	2.13	2.33	2.25	2.38	2.31	3.20	2.36	2.41	2.24	2.50	2.57	2.81	2.84	4.45		
Q23	1.18	1.32	1.33	1.42	1.37	1.49	1.37	1.27	1.18	1.24	1.46	1.61	1.44	1.72	1.95	
Q24	1.98	2.35	2.12	2.28	2.44	3.25	2.46	2.42	2.35	2.59	2.23	2.38	2.42	3.63	1.52	4.65

ML with missing data

	Q1	Q3	Q4	Q6	Q7	Q8	Q10	Q15	Q16	Q17	Q18	Q19	Q21	Q22	Q23	Q24
Q1	2.69															
Q3	2.11	2.89														
Q4	1.79	1.92	2.79													
Q6	1.99	2.19	2.46	2.88												
Q7	1.93	2.18	1.82	2.09	3.55											
Q8	2.53	2.95	2.37	2.76	3.11	6.86										
Q10	1.89	2.19	1.80	2.11	2.86	3.04	3.54									
Q15	1.95	2.32	1.93	2.10	2.21	2.98	2.12	2.75								
Q16	1.81	1.90	1.66	1.76	1.99	2.84	1.97	2.18	3.34							
Q17	1.72	2.00	1.65	1.87	2.11	2.80	2.23	2.23	2.05	3.56						
Q18	1.66	1.92	1.54	1.81	1.90	2.50	1.74	1.97	2.09	2.03	4.47					
Q19	1.87	2.16	1.77	2.03	2.15	2.69	2.13	2.22	2.14	2.21	2.93	3.62				
Q21	1.72	1.98	1.73	1.87	1.88	2.53	1.96	1.89	1.78	1.88	1.78	2.09	3.14			
Q22	2.12	2.30	2.22	2.39	2.31	3.24	2.39	2.36	2.22	2.44	2.54	2.88	2.83	4.45		
Q23	1.20	1.33	1.33	1.44	1.42	1.60	1.40	1.27	1.21	1.25	1.44	1.56	1.45	1.72	1.95	
Q24	1.90	2.26	2.05	2.22	2.48	3.19	2.42	2.35	2.40	2.62	2.30	2.44	2.37	3.59	1.55	4.66

Mean imputation

	Q1	Q3	Q4	Q6	Q7	Q8	Q10	Q15	Q16	Q17	Q18	Q19	Q21	Q22	Q23	Q24
Q1	2.64															
Q3	2.03	2.87														
Q4	1.74	1.91	2.78													
Q6	1.93	2.18	2.45	2.87												
Q7	1.84	2.09	1.74	1.98	3.44											
Q8	2.16	2.51	2.06	2.38	2.71	6.08										
Q10	1.73	2.09	1.73	2.01	2.68	2.54	3.41									
Q15	1.91	2.30	1.93	2.08	2.18	2.66	2.05	2.76								
Q16	1.72	1.81	1.59	1.68	1.93	2.48	1.85	2.12	3.22							
Q17	1.56	1.83	1.54	1.74	1.94	2.48	2.02	2.06	1.85	3.25						
Q18	1.57	1.82	1.51	1.77	1.73	2.12	1.56	1.86	1.83	1.84	4.19					
Q19	1.76	1.99	1.71	1.91	1.96	2.27	1.95	2.09	1.97	1.97	2.60	3.40				
Q21	1.66	1.93	1.70	1.82	1.81	2.22	1.89	1.85	1.68	1.75	1.68	1.97	3.09			
Q22	2.05	2.26	2.15	2.30	2.21	2.84	2.23	2.32	2.14	2.30	2.45	2.61	2.72	4.33		
Q23	1.20	1.33	1.32	1.44	1.40	1.35	1.40	1.27	1.18	1.18	1.43	1.55	1.41	1.67	1.95	
Q24	1.71	2.06	1.82	1.97	2.23	2.54	2.11	2.16	2.19	2.24	1.98	2.09	2.08	3.21	1.46	4.27

Table 4: Covariance matrices obtained under 4 alternative missing data treatments

The results of the final CFA model are compared for the seven estimation methods. Table 4 shows the estimated sample covariances. It can easily be seen that covariances obtained under imputation or ML are very similar. This argues for the quality of the imputation, though it can also be a result of the low rate of data missingness. On the contrary, variances and covariances resulting from mean imputation are in some cases considerably lower, especially when covariances involve pairs of variables with a sizeable amount of missing data. In SEM, biased covariances can only lead to biased parameter estimates.

Table 5 shows the estimates and goodness of fit statistics of the CFA model shown in the path diagram of Figure 2 using the seven methods. Several goodness of fit measures are usually considered in SEM (Bollen & Long, 1993). A likelihood ratio χ^2 test of the hypothesis that all model constraints hold in the population is usually performed first. Here we notice the first differences across methods. All procedures that are not robust to non-normality clearly lead to the rejection of the model's constraints. For ML with robust test statistics, the hypothesis is also rejected but by a narrower margin. Usually researchers are not so interested in exactly fitting models, so that quantitative measures of misfit are preferred to tests of exact fit. A wealth of fit measures have been suggested. Among the most widely used ones are the Root Mean Squared Error of Approximation (RMSEA), the Tucker and Lewis Index (TLI), also known as Non Normed Fit Index, and the root mean squared residual correlation (RMSR). RMSEA and TLI take the parsimony of the model into account, so that the releasing of approximately correct constraints does not necessarily improve the values of these indices. Values of RMSEA and RMSR below 0.05 and values of TLI above 0.95 are usually considered acceptable, though the debate concerning which goodness of fit measures to use and what the threshold for a good model can be is far from resolved (see Bollen & Long, 1993 for details). Using these measures, the model looks acceptable, but even more so when robust methods are considered. None of these fit indices are available for PLS. PLS focuses only on predictive fit measures that, except for the error variances and the standardized loadings, are irrelevant with the purpose of validating a questionnaire using a CFA model. The next part of the table shows the point estimates (raw and standardized; they are identical for robust and non-robust methods using the same missing value treatment) and the standard errors (robust and standard ML). Three types of parameters are present: factor loadings, factor correlations and error variances. Standardized error variances are equal to one minus the R-square of each measurement equation. Within each parameter type, averages of the relevant statistics are also shown.

The point estimates obtained from imputed data or by direct ML with missing data are generally very similar. This could be expected from the fact that covariances are also similar. The only exception to this is the factor correlations. Simple imputation methods such as hot deck are known to inflate correlations. However, on average the differences are not dramatic. Overall, it looks as if there is not much to be gained by using direct ML with missing data when the proportion of data that are missing is low. Conversely, there seems to be a great deal to be lost out of using obsolete missing data treatments such as mean imputation. Nearly all unstandardized factor loadings are lower than for any other of the three methods, and nearly all unstandardized error variances are higher. The differences are larger for variables with a larger proportion of missing cases.

Non robust ML standard errors are known to underestimate true standard errors when most variables have a positive kurtosis. This seems to be the case for our data set and for all missing value treatments, as robust standard errors are higher in nearly all cases, for some parameters more than twice as high. The failure to use robust standard errors may thus result in confidence intervals that do not contain the true parameter value or in inflated t-values rejecting true null hypotheses. The advocates of PLS could not be more right in claiming that non-normality was a key issue in SEM.

As regards robust standard errors, they are relatively similar across missing value treatments. It is known that imputation leads to an underestimation of uncertainty, regardless of whether it is done with mean substitution or hot deck methods (only multiple imputation is immune to this). This is also observed here, although differences are not very large (nearly zero for factor loadings and in the neighbourhood of 10% for the remaining parameter types). The sample size drops from 258 to 250 for the hot deck case, but the effect of such a small sample size change on the above results would hardly go noticed, as the root of the sample size ratio is only 1.015; thus accounting for only 1.5% difference in standard errors.

For our data set, that contains a small amount of missing data but severely non-normal data, the use of an imputation method and an estimation method with robust standard errors and test statistics seems to be a wise second-best option for those researchers who do not have the latest software available.

As regards PLS, its results can be compared to the SEM results using the same data, that is, the mean imputed data, and tests robust to non-normality. We find that loadings are systematically higher for PLS and the factor correlations and the error variances systematically lower. Besides, loadings on the same factor are more similar under PLS than under SEM. The differences between SEM and PLS are higher than between any pair of SEM methods, and can be attributed to the well known fact that PLS is only consistent for an infinite number of indicators per latent variable, as non-normality does not affect the consistency of SEM estimates (Satorra 1990, 1992, 1993, Browne, 1984; Fouladi, 2000; Nevitt & Hancock, 2001). When validating a questionnaire using a CFA model, PLS can result in retaining items with an apparently high but actual low loading and can also result in related factors appearing unrelated. This result can be understood because for CFA models, PLS yields exactly the same estimates as principal component analyses done separately for each dimension. Standard errors cannot be compared to SEM standard errors, as different estimation procedures can also differ in precision. They can anyway be trusted as the jackknife procedure is robust to non normality. However, one may wonder about the use of getting the correct standard errors for the wrong point estimates. PLS were also used on imputed data and compared to SEM on imputed data with robust standard errors with identical conclusions. The results are available from the authors on request.

	ML with missing data				ML on imputed data				ML on mean substituted data				PLS on mean subst.		
Test statistics			robust	ML			robust	ML			robust	ML			
χ^2			125.51	153.04			58.80	157.83			62.78	166.15			
d.f.			89	89			42	89			43	89			
p-value			0.007	0.000			0.044	0.000			0.026	0.000			
RMSEA			0.040	0.053			0.040	0.056			0.042	0.058			
SRMR			0.029	0.029			0.029	0.029			0.029	0.029			
TLI			0.978	0.975			0.982	0.973			0.978	0.969			
	estim	stand. estim.	robust s. e.	ML s. e.	estim	stand. estim.	robust s. e.	ML s. e.	estim	stand. estim.	robust s. e.	ML s. e.	estim	stand. estim.	robust s. e.
Loadings															
Q1 on QUAL1	1.34	0.82	0.091	0.085	1.34	0.82	0.090	0.086	1.31	0.81	0.089	0.085	1.26	0.78	0.100
Q3 on QUAL1	1.53	0.90	0.091	0.084	1.53	0.90	0.092	0.085	1.50	0.89	0.089	0.084	1.40	0.83	0.120
Q8 on QUAL1	1.98	0.76	0.130	0.147	1.93	0.75	0.129	0.140	1.73	0.70	0.133	0.136	2.29	0.91	0.130
Q4 on QUAL2	1.48	0.89	0.103	0.083	1.49	0.89	0.105	0.085	1.48	0.89	0.102	0.082	1.61	0.97	0.080
Q6 on QUAL2	1.66	0.98	0.096	0.079	1.66	0.98	0.096	0.080	1.65	0.98	0.095	0.079	1.64	0.97	0.080
Q7 on VALU	1.70	0.90	0.100	0.096	1.63	0.89	0.101	0.094	1.65	0.89	0.100	0.094	1.75	0.94	0.090
Q10 on VALU	1.69	0.90	0.089	0.096	1.63	0.87	0.090	0.097	1.61	0.88	0.091	0.094	1.74	0.94	0.070
Q15 on SATI	1.55	0.93	0.096	0.080	1.55	0.93	0.099	0.081	1.54	0.93	0.096	0.080	1.50	0.90	0.090
Q16 on SATI	1.40	0.77	0.101	0.099	1.38	0.78	0.100	0.095	1.36	0.76	0.100	0.096	1.56	0.87	0.100
Q17 on SATI	1.44	0.77	0.118	0.105	1.44	0.77	0.115	0.101	1.35	0.75	0.117	0.097	1.55	0.86	0.100
Q18 on LOYA	1.63	0.77	0.149	0.119	1.56	0.74	0.145	0.118	1.54	0.75	0.145	0.114	1.92	0.94	0.130
Q19 on LOYA	1.81	0.95	0.129	0.098	1.73	0.92	0.135	0.100	1.69	0.92	0.131	0.096	1.65	0.90	0.160
Q21 on IMAG	1.44	0.82	0.130	0.093	1.45	0.82	0.131	0.094	1.42	0.81	0.130	0.092	1.47	0.84	0.130
Q22 on IMAG	1.92	0.91	0.114	0.104	1.93	0.92	0.116	0.104	1.86	0.90	0.119	0.103	1.95	0.94	0.090
Q23 on IMAG	0.96	0.68	0.122	0.079	0.94	0.68	0.127	0.080	0.97	0.69	0.126	0.078	0.97	0.69	0.150
Q24 on IMAG	1.79	0.82	0.119	0.117	1.79	0.83	0.112	0.113	1.61	0.78	0.124	0.110	1.81	0.88	0.100
Average	1.58	0.85	0.111	0.098	1.56	0.84	0.111	0.097	1.52	0.83	0.112	0.095	1.63	0.88	0.108
Factor correlations															
QUAL1withQUAL2	0.87	0.87	0.036	0.023	0.86	0.86	0.035	0.024	0.87	0.87	0.035	0.023	0.72	0.72	0.045
QUAL1 with VALU	0.86	0.86	0.029	0.027	0.88	0.88	0.027	0.026	0.86	0.86	0.028	0.028	0.73	0.73	0.039
QUAL1 with SATI	0.95	0.95	0.020	0.017	0.96	0.96	0.016	0.016	0.96	0.96	0.019	0.017	0.79	0.79	0.036
QUAL1 with LOYA	0.77	0.77	0.060	0.036	0.81	0.81	0.051	0.034	0.79	0.79	0.055	0.037	0.62	0.62	0.047
QUAL1 with IMAG	0.84	0.84	0.034	0.027	0.84	0.84	0.031	0.027	0.85	0.85	0.030	0.027	0.72	0.72	0.043
QUAL2 with VALU	0.74	0.74	0.042	0.034	0.76	0.76	0.041	0.033	0.73	0.73	0.042	0.035	0.66	0.66	0.036
QUAL2 with SATI	0.81	0.81	0.045	0.028	0.82	0.82	0.040	0.027	0.80	0.80	0.041	0.028	0.70	0.70	0.041
QUAL2 with LOYA	0.68	0.68	0.068	0.041	0.72	0.72	0.063	0.039	0.68	0.68	0.065	0.041	0.59	0.59	0.056
QUAL2 with IMAG	0.77	0.77	0.045	0.031	0.77	0.77	0.043	0.031	0.77	0.77	0.042	0.031	0.71	0.71	0.042
VALU with SATI	0.84	0.84	0.030	0.028	0.86	0.86	0.029	0.026	0.85	0.85	0.029	0.027	0.74	0.74	0.037
VALU with LOYA	0.70	0.70	0.065	0.042	0.74	0.74	0.056	0.041	0.70	0.70	0.059	0.043	0.57	0.57	0.053
VALU with IMAG	0.77	0.77	0.039	0.034	0.79	0.79	0.037	0.033	0.78	0.78	0.038	0.034	0.69	0.69	0.038
SATI with LOYA	0.81	0.81	0.064	0.033	0.83	0.83	0.055	0.032	0.82	0.82	0.057	0.033	0.69	0.69	0.035
SATI with IMAG	0.84	0.84	0.037	0.026	0.85	0.85	0.033	0.025	0.85	0.85	0.034	0.026	0.77	0.77	0.033
LOYA with IMAG	0.82	0.82	0.050	0.031	0.84	0.84	0.049	0.031	0.83	0.83	0.050	0.032	0.69	0.69	0.054
Average	0.80	0.80	0.044	0.031	0.82	0.82	0.040	0.030	0.81	0.81	0.042	0.031	0.69	0.69	0.042
Error Variances															
Q1	0.88	0.33	0.313	0.090	0.90	0.33	0.289	0.092	0.92	0.35	0.276	0.093	1.03	0.39	
Q3	0.55	0.19	0.082	0.072	0.55	0.19	0.078	0.071	0.61	0.21	0.088	0.076	0.91	0.32	
Q4	0.59	0.21	0.091	0.074	0.60	0.21	0.089	0.075	0.58	0.21	0.090	0.073	1.02	0.16	
Q6	0.13	0.04	0.074	0.066	0.12	0.04	0.072	0.066	0.14	0.05	0.072	0.066	0.19	0.07	
Q7	0.67	0.19	0.126	0.111	0.69	0.21	0.114	0.108	0.70	0.20	0.118	0.112	0.18	0.06	
Q8	2.96	0.43	0.564	0.301	2.87	0.43	0.509	0.278	3.06	0.51	0.474	0.286	0.37	0.11	
Q10	0.69	0.20	0.122	0.112	0.82	0.24	0.116	0.115	0.79	0.23	0.114	0.113	0.37	0.11	
Q15	0.36	0.13	0.071	0.063	0.37	0.13	0.064	0.061	0.38	0.14	0.068	0.063	0.51	0.18	
Q16	1.36	0.41	0.255	0.135	1.21	0.39	0.166	0.120	1.36	0.42	0.191	0.131	0.78	0.24	
Q17	1.47	0.41	0.186	0.150	1.41	0.40	0.170	0.138	1.43	0.44	0.154	0.137	0.83	0.26	
Q18	1.83	0.41	0.580	0.203	1.95	0.45	0.553	0.208	1.82	0.44	0.528	0.195	0.49	0.12	
Q19	0.36	0.10	0.205	0.146	0.58	0.16	0.218	0.149	0.54	0.16	0.195	0.143	0.67	0.20	
Q21	1.05	0.33	0.195	0.110	1.04	0.33	0.180	0.110	1.06	0.35	0.176	0.112	0.92	0.30	
Q22	0.74	0.17	0.141	0.111	0.72	0.16	0.131	0.107	0.84	0.20	0.135	0.116	0.52	0.12	
Q23	1.04	0.53	0.182	0.099	1.06	0.55	0.157	0.101	1.01	0.52	0.147	0.096	1.01	0.52	
Q24	1.53	0.32	0.227	0.171	1.42	0.31	0.178	0.154	1.65	0.39	0.202	0.168	0.98	0.23	
Average	1.01	0.28	0.213	0.126	1.02	0.28	0.193	0.122	1.06	0.27	0.189	0.124	0.67	0.21	

Table 5: Estimates of the CFA model using seven different methods

The goodness of fit of the model can be interpreted in terms of measurement validity. The constraints that are being tested by the χ^2 test in a CFA model are the absence of loadings on more than one factor and the absence of error covariances (extraneous sources of common variance) which, if present, would reveal invalidity of the items involved. The standardized loadings can be interpreted as indicators of reliability of each of the items. All are around or above 0.70.

Factor correlations between the latent variables are very high. This can lead to collinearity problems making it difficult to identify the significant causal relationships between the latent variables. Besides, some of the correlations are very close to one; for instance, the correlation between QUAL1 and satisfaction, which, fortunately, is still significantly different from one, though by a narrow margin. A correlation equal to one between two concepts would show that respondents do not distinguish between them. If PLS had been considered, the closeness of this correlation to unity would have gone unnoticed.

3.2. Complete Model

The CFA model was reparametrized into a complete SEM specifying regression equations among factors. This model was further modified to improve its fit to the data.

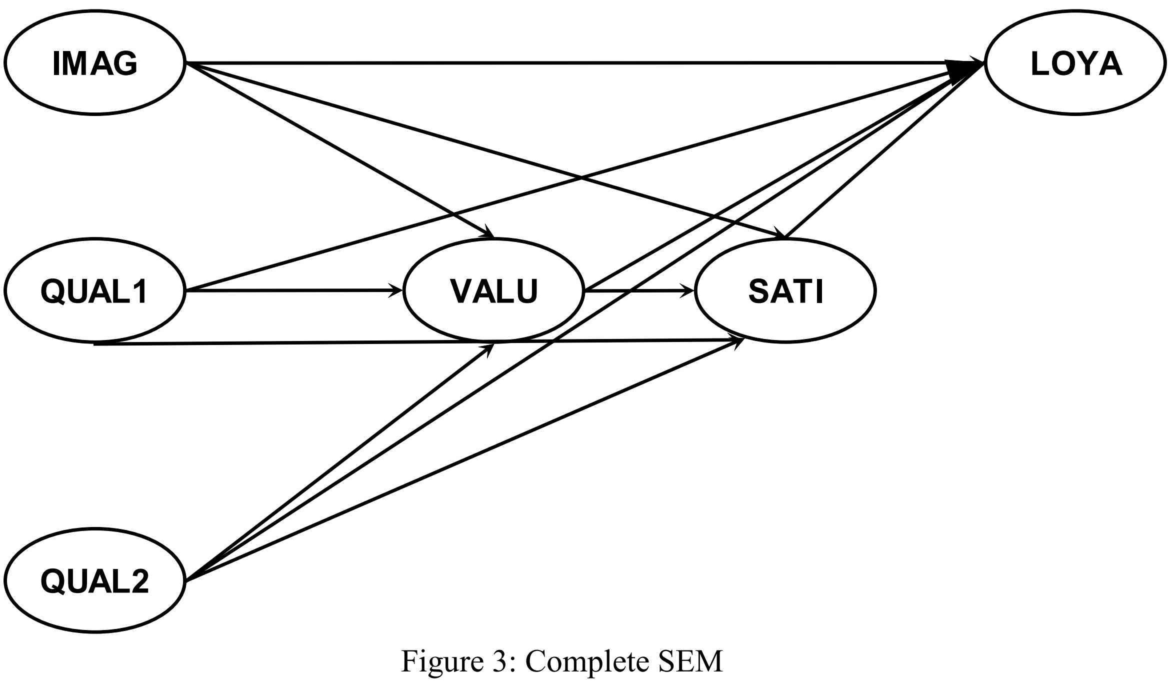

Figure 3: Complete SEM

Table 6 shows the results of the first, last and some intermediate steps of the model modification process. For simplicity, only t-values and standardized estimates of parameters in the equations relating factors to one another (the so-called structural part of the model) are shown. Robust direct ML estimation with missing data was used throughout.

	Model 1		Model 2		Model 3		Model 4	
χ^2		125.51		126.42		127.51		130.96
d.f.		89		92		94		97
p-value		0.007		0.010		0.012		0.012
scaling constant c		1.219		1.214		1.215		1.223
Robust χ^2 difference test			0.45		1.15		3.55	
d.f. difference			3		2		3	
p-value			0.929		0.563		0.314	
RMSEA		0.040		0.038		0.037		0.037
SRMR		0.029		0.029		0.030		0.032
TLI		0.978		0.980		0.981		0.981
	t-value	stand. estim.	t-value	stand. estim.	t-value	stand. estim.	t-value	stand. estim.
Regression coefficients								
VALU on IMAG	1.37	0.17	1.40	0.17	1.51	0.18		
VALU on QUAL1	4.20	0.73	4.24	0.73	5.98	0.70	12.80	0.87
VALU on QUAL2	-0.18	-0.02	-0.17	-0.02				
SATI on VALU	0.45	0.05	0.44	0.05	0.75	0.07		
SATI on IMAG	1.58	0.16	1.63	0.17	1.68	0.17		
SATI on QUAL1	4.31	0.87	4.35	0.86	5.45	0.74	16.80	0.95
SATI on QUAL2	-0.86	-0.10	-0.89	-0.11				
LOYA on SATI	1.60	0.51	2.51	0.39	2.53	0.39	2.71	0.38
LOYA on VALU	-0.03	-0.00						
LOYA on IMAG	3.01	0.52	2.95	0.49	2.99	0.49	3.32	0.50
LOYA on QUAL1	-0.33	-0.12						
LOYA on QUAL2	-0.16	-0.02						
Covariances among exogenous variables								
QUAL1 with QUAL2	7.44	0.87	7.43	0.87	7.34	0.86	7.29	0.86
QUAL1 with IMAG	6.46	0.84	6.46	0.84	6.47	0.84	6.56	0.86
QUAL2 with IMAG	5.97	0.77	5.95	0.77	5.96	0.77	5.97	0.77
R^2								
VALU		0.74		0.74		0.74		0.76
SATI		0.92		0.91		0.90		0.91
LOYA		0.72		0.71		0.71		0.71

Table 6: Estimates of the structural part of 4 nested models.
Robust direct ML with imputed data.

First, a saturated structural model was built assuming only the causal ordering of the variables most often encountered in the literature (see Figure 1). All possible parameters fitting within this causal order were freely estimated (see Figure 3). Effects are assumed to flow from quality and image to value, satisfaction and loyalty. All covariances among the three exogenous factors are free parameters to be estimated. The model is equivalent to the CFA model and has the same goodness of fit statistics.

This first model is displayed in the first column of Table 6. Many of the structural parameter coefficients were insignificant and were removed from the model one at a time to increase parsimony and identify changes in the significance of parameters that may have been previously hidden because of collinearity Robust χ^2 difference tests for certain nested models were computed. Additionally, other goodness of fit indices are used, in particular those that take parsimony into account (i.e. RMSEA and TLI).

Parameters that are both theoretically and statistically insignificant (i.e., those with t-values lower than 2 in absolute value and not present in the path diagram of Figure 1) were removed first, starting with the ones with the lowest t-values. This led to removing one by one and in this order the regression coefficients of loyalty on value, loyalty on quality2 (service), and loyalty on quality1 (product). At all steps, fit measures that take parsimony into account showed no deterioration and even slight improvements. The modified model is displayed in the second column of Table 6. The regression coefficient of loyalty on satisfaction became significant during these modifications. The significance of all other parameters and the R^2 hardly changed. An overall χ^2 difference test of this model against the previous unrestricted one also leads to maintaining the restrictions introduced so far.

In this model it was observed that quality2 had two non-sensically negative effects upon value and satisfaction that were statistically insignificant. They were removed in turn, which led to the model in the third column of Table 6. Once more, R^2 and global fit measures that take parsimony into account showed no deterioration and a joint test of both restrictions using the χ^2 difference test led to maintaining the restricted model. This suggests that once product quality has been taken into account, service quality does not add any extra explanatory power.

All non-significant coefficients in the model of the third column of Table 6 have the expected sign and are theoretically relevant. Many researchers are reluctant to drop relevant coefficients, therefore many readers will feel comfortable with ending the modification process here and interpreting this model.

As an alternative, we can continue to remove non-significant parameters one by one, starting by those with the lowest t-values; first the coefficient of satisfaction on value, then of value on image and finally of satisfaction on image. The final model is displayed in the last column of Table 6. Once more, R^2 and global fit measures that take parsimony into account showed no deterioration and a joint test of both restrictions using the χ^2 difference test led to maintaining the restricted model.

The final model indicates that the main and only driver of consumer satisfaction is the quality of the Post Office products. Neither service quality nor value, nor image have a significant impact on satisfaction levels. In fact, the quality of the Post Office customer service or “humanware” appears to have no significant impact on any of the factors in the model. Thus, similar to Post Denmark (Kristensen et al., 1999) the delineation between service and product quality is necessary.

As expected, both image and satisfaction are important factors in determining customer loyalty with image being the most important predictor of customer loyalty. The significant impact of image in generating customer loyalty was also noted in the Post Denmark study.

The significance of the effect of quality1 on value is not unexpected given the inextricable relationship between both as well as the wording of the items in the value factor (quality given price and price given quality). Other effects on value present in Figure 1 (quality2 and image) were non significant. In fact, the effect of image on value is not always specified. These results also argue for a differential behaviour of both types of quality.

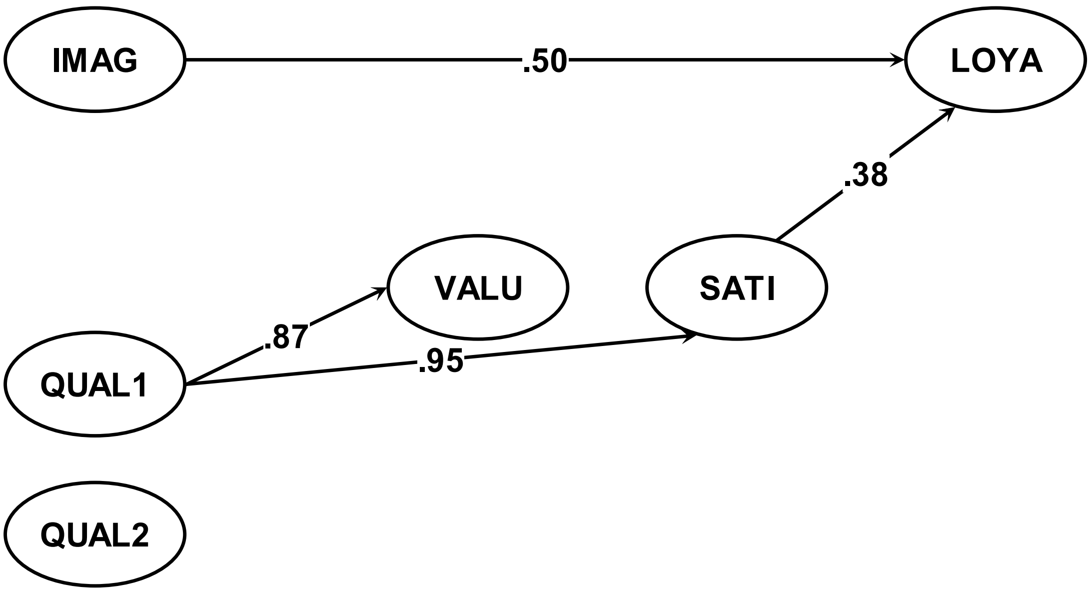

Figure 4: Final SEM

4. Discussion

From a methodological point of view, the paper has shown that proper treatments of missing data and non-normality are important when modelling the ESCI/ASCI and that, unlike the case was previously, these treatments are now available for SEM and offer some advantages over the competing PLS methodology. The most important advantage is that SEM estimates are free of bias, which has serious implications when validating a questionnaire and when estimating relationships among factors.

From a substantive point of view, the fact that product quality turned out to be such an important factor in the model is not surprising from its correlation with satisfaction as shown in the CFA models. That is why product quality seems to be the only variable required to explain and predict satisfaction. From an examination of the wording of the items it is hard to believe that both factors actually measure the same thing therefore we are inclined to maintain discriminant validity as their correlation is significantly different from 1, even if only by a narrow margin. There is ample evidence in the literature suggesting that perceived quality is the most significant predictor of satisfaction (e.g., Anderson & Sullivan, 1993; Churchill & Suprenant, 1982; Johnson & Fornell, 1991). However, Martensen et al., (2000) noted that the drivers of both customer satisfaction and loyalty are in fact industry specific. For example they demonstrated that image was the main driver of customer satisfaction in industries like cable Television, supermarkets and the internet. In contrast, satisfaction levels in the soft drinks industry were driven by product quality. The distinction between service quality and product quality proved justified in this analysis. In fact, service quality does not have a significant impact on value, satisfaction or loyalty. This result may be specific to postal services or may have been caused by the high collinearity between product and service quality.

A very interesting, though not surprising result was the impact of image on loyalty. In fact image is the most important predictor of loyalty to the Isle of Man Post Office. Satisfaction predicted loyalty to a lesser degree. A similar result was documented by Kristensen et al., (1999) in relation to Post Denmark where image was found to be the most significant predictor of loyalty. Indeed this is a very important observation as competition within this segment of the market is going to increase in the future. Martensen et al., (2000) noted that in the fast food, internet and soft drinks industries the main impact on loyalty comes from the product itself (product quality as opposed to service quality), whereas in more complex industries with more highly competitive markets (e.g., mobile phone industries) and industries with multiple outlets (e.g., supermarkets and Banks) loyalty is more image driven. Thus, the drivers of customer loyalty are in fact industry specific. In this analysis both product and service quality do not have a significant impact on loyalty. Martensen et al., (2000) argued that loyalty was in fact the most significant outcome measure in the CSI model and was a more *true* measure of reality compared to satisfaction. They maintained that to recommend a product or service to others has greater consequences and requires more commitment than just indicating that one is more or less satisfied compared to an ideal or a competitive product or service.

In the CSI literature a more complex relationship between quality, value and satisfaction is espoused. This relationship was partly realised in this analysis with only product quality having a significant impact on value. As quality is a major aspect of the value assessment, and given the definition of value this result is not surprising. However, because of the definitional interaction between quality and value Johnson et al., (2001) maintain that it is difficult to evaluate how much of the impact of quality on value is due to cause and effect and how much is true by definition. As such they suggested replacing this factor in the model with a perceived price construct with customers evaluating price relative to a variety of benchmarks, including comparisons of the product's price versus expected price, competitors prices and quality, thus producing a more *pure* price construct. In fact, the focus groups suggested that the wording of items in the value construct were confusing.

The fact that all factors are so highly correlated with all others raise two further points. Firstly, in a study like ours, where data are measurements obtained in a single questionnaire, it is possible that correlations are artifactually inflated by what is known as common method variance (See Linden & Whitney, 2001; Harrison et al., 1996). Secondly, statistically, there is an alternative way of structuring the relationships among factors when all factor correlations are high. That is to specify a higher order factor structure between all the factors in the model. This model examines the communalites between all factors implying that all are in fact measuring a global conceptualisation of consumer satisfaction. The intricate causal relationships (both direct and indirect) espoused in the ECSI/ACSI literature would thus be ignored. The fit of such a model to our data was nearly equally good to the fit of the model specifying causal relationships among factors.

From a statistical point of view, using the data to modify the specification of the model, as we have done, makes the results vulnerable to capitalization on chance (Luijben, 1989; MacCallum, 1986; MacCallum, 1995; MacCallum et al., 1992), a risk that can only increase when collinearity is high. Some of the recommendations of the literature to prevent capitalization on chance (make only one modification at a time, use both the overall fit of the model and the significance of the individual parameters, make theoretically sound modifications first) could be followed. Others could not. Our sample was not large enough to spare a part of it for a crossvalidation exercise. However, some results hint at the fact that the model modification was successful. We repeated the modification several times by removing different effects on the first step and the final model was always the same. Along all this modification process, only one parameter shifted from non significant to significant (the effect of satisfaction on loyalty) all the rest were either always significant or always non-significant.

References

Aburckle, J. L. (1996), Full Information Estimation in the Presence of Incomplete Data, in Marcoulides, G.A. and Schumacker, R.E. (eds.), *Advanced Structural Equation Modeling,* Mahwah, NJ, Erlbaum, (pp. 243-277).

Aburckle, J. L. and Wothke, W. (1999), *Amos (Version 4.0),* Chicago, IL, Smallwaters.

Allison, P. D. (1987), Estimation of Linear Models with Incomplete Data, in Clogg, C.C. (ed.), *Sociological Methodology, 1987,* Washington DC, American Sociological Association, (pp. 71-103).

American Society for Quality (1998), *American Customer Satisfaction Index (ACSI) Methodology Report*, Ann Arbor, Mi: Arthur Andersen, University of Michigan.

Anderson, E.W. and Fornell, C. (2000), Foundations of the American Customer Satisfaction Index, *Total Quality Management* 11, S869-S882.

Anderson, E.W., Fornell, C. and Lehrmann, D.R. (1994), Customer Satisfaction, Market Share, and Profitability: Findings from Sweden. *Journal of Marketing* 58, 53-66.

Anderson, E. W. and Sullivan, M. (1993), The Antecedents and Consequences of Customer Satisfaction for Firms, *Marketing Science* 12, 125-143.

Andreassen, T. W. and Lindestad, B. (1998a), The Effects of Corporate Image in the Formation of Customer Loyalty, *Journal of Service Marketing* 1, 82-92.

Andreassen, T. W. and Lindestad, B. (1998b), Customer Loyalty and Complex Services, the Impact of Corporate Image on Quality, Customer Satisfaction and Loyalty for Customers with Varying Degrees of Service Expertise, *International Journal of Service Industry Management* 9, 7-23.

Arminger, G. and Sobel, M. E. (1990), Pseudo Maximum Likelihood Estimation of Mean and Covariance Structures with Missing Data, *Journal of the American Statistical Association* 85, 195-203.

Barnett, V. and Lewis, T. (1994), *Outliers in Statistical Data,* New York, Wiley.

Batista-Foguet, J. M. and Coenders, G. (1998), Introducción a los Modelos Estructurales. Utilización del Análisis Factorial Confirmatorio para la Depuración de un Cuestionario [Introduction to Structural Equation Models. Using a Confirmatory Factor Analysis Model to Improve a Questionnaire], in Renom, J. (ed.) *Tratamiento Informatizado de Datos [Computer Data Analysis]*, Barcelona, Masson, (pp. 229-286).

Batista-Foguet, J. M. and Coenders, G. (2000), *Modelos de Ecuaciones Estructurales [Structural Equation Models]*, Madrid, La Muralla.

Bentler, P.M. (2000), *EQS 6.0,* Encino, CA, *Multivariate Software*.

Bollen, K. A. (1989), *Structural Equations with Latent Variables*, New York, Wiley.

Bollen, K.A. and Long, J.S. (1993), *Testing Structural Equation Models.* Newbury Park, Ca, Sage.

Browne, M.W. (1984), Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures, *British Journal of Mathematical and Statistical Psychology* 37, 62-83.

Cassel, C., Hackl, P. and Westlund, A.H. (2000), On Measurement of Tangible Assets, a Study of Robustness of Partial Least Squares, *Total Quality Management* 11, 897-907.

Churchill, G. A., and Suprenant, C. (1982), An Investigation Into the Determinants of Consumer Satisfaction, *Journal of Marketing Research* 19, 491-504.

Coenders, G. and Saris, W. E. (1995), Categorization and Measurement Quality. The Choice Between Pearson and Polychoric Correlations, in Saris, W.E. and Münnich, Á., (eds.), *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments*, Budapest, Eötvös University Press, (pp. 125-144).

Coenders, G., and Saris, W. E. (2000), Testing Nested Additive, Multiplicative and General Multitrait-Multimethod Models, *Structural Equation Modeling* 7, 219-250.

Coenders, G., Satorra, A. and Saris, W. E. (1997), Alternative Approaches To Structural Modeling of Ordinal Data, a Monte Carlo Study, *Structural Equation Modeling* 4, 261-282.

Dijkstra, T. (1983), Some Comments on Maximum Likelihood and Partial Least Squares Methods, *Journal of Econometrics* 22, 67-90.

ECSI Technical Committee (1998), *European Customer Satisfaction Index, Foundation and Structure for Harmonized National Pilot Projects.* Report Prepared for the ECSI Steering Committee, October.

Eklöf, J. A. (2000), *European Customer Satisfaction Index Pan-European Telecommunication Sector Report Based on the Pilot Studies 1999.* Stockholm, Sweden, European Organization for Quality and European Foundation for Quality Management.

Enders, C.K. (2001), The Impact of Non-Normality on Full Information Maximum Likelihood Estimation for Structural Equation Models with Missing Data, *Psychological Methods* 6, 352-370.

Enders, C.K. and Bandalos, D.L. (2001), The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models, *Structural Equation Modeling* 8, 430-457.

Fornell, C. (1992), A National Customer Satisfaction Barometer, the Swedish Experience, *Journal of Marketing* 56, 6-21.

Fornell, C. and Cha, J. (1994), Partial Least Squares, in Bagozzi, R.P., (ed.), *Advanced Methods in Marketing Research,* Cambridge, Basil Blackwell, (pp. 52-78).

Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., and Bryant, B. E. (1996), The American Customer Satisfaction Index, Nature, Purpose and Findings, *Journal of Marketing* 60, 7-18.

Fouladi, R.T. (2000), Performance of Modified Test Statistics in Covariance and Correlation Structure Analysis under Conditions of Multivariate Non-Normality, *Structural Equation Modeling* 7, 356-410.

Graham, J.W., Hofer, S.M. and Mackinnon, D.P. (1996), Maximizing the Usefulness of Data Obtained with Planned Missing Data Patterns, an Application of Maximum Likelihood Procedures, *Multivariate Behavioral Research* 31, 197-218.

Graham, J.W., Hofer, S.M. and Piccinin, A.M. (1994), Analysis with Missing Data in Drug Prevention Research, in Collins, L.M. and Seitz, L., (eds.), *Advances in Data Analysis for Prevention Intervention Research,* Washington, DC, National Institute on Drug Abuse, (pp. 13-63).

Graham, J.W., Taylor, B.J. and Cumsille, P.E. (2001), Planned Missing Data Designs in the Analysis of Change, in Collins, L.M. and Sayer, A., (eds.), *New Methods for the Analysis of Change,* Washington, DC, APA., (pp. 333-354).

Groves, R. M. (1989), *Survey Errors and Survey Costs.* New York: John Wiley & Sons.

Hackl, P., Scharitzer, D. and Zuba, R. (2000), Customer Satisfaction in the Austrian Food Retail Market, *Total Quality Management* 11, 999-1006.

Harrison, D. A., McLaughlin, M. E. and Coalter, T. M. (1996), Context, Cognition, and Common Method Variance: Psychometric and Verbal Protocol Evidence, *Organizational Behavior and Human Decision Processes* 68, 246-261.

Howard, J. A. and Sheth, J. N. (1969), The *Theory of Buyer Behavior,* New York, Wiley.

Johnson, M.D. and Fornell, C. (1991), A Freamework for Comparing Customer Satisfaction across Individuals and Product Categories, *Journal of Economic Psychology* 12, 267-286.

Johnson, M. D., Gustafsson, A., Andreassen, T. W., Lervik, L. and Cha, J. (2001), The Evolution and Future of National Customer Satisfaction Index Models, *Journal of Economic Psychology* 22, 217-245.

Jöreskog, K. G. (1973), A General Method for Estimating a Linear Structural Equation System, in Goldberger, A.S. and Duncan, O.D. (eds.), *Structural Equation Models in the Social Sciences,* New York, Academic Press, (pp. 85-112).

Jöreskog, K. G. (1990), New Developments in LISREL. Analysis of Ordinal Variables Using Polychoric Correlations and Weighted Least Squares, *Quality and Quantity* 24, 387-404.

Jöreskog, K. G. and Sörbom, D. (1993), *New Features in PRELIS2*, Chicago, IL, Scientific Software International.

Jöreskog, K. G, Sörbom, D., Du Toit, S., and Du Toit, M. (2000), *LISREL 8, New Statistical Features (Rev.),* Chicago, IL, Scientific Software International.

Kristensen, K., Martensen, A. and Grønholdt, L. (1999), Measuring the Impact of Buying Behaviour on Customer Satisfaction, *Total Quality Management* 10, 602-614.

Linden, M. K. and Whitney, D. J. (2001), Accounting for Common Method Variance in Cross-Sectional Research Designs, *Journal of Applied Psych* 86, 114-121.

Little, R.J.A. and Rubin, D.B. (1987), *Statistical Analysis with Missing Data,* New York, Wiley.

Luijben, T. (1989), *Statistical Guidance for Model Modification in Covariance Structure Analysis,* Amsterdam, Sociometric Research Foundation.

Maccallum, R.C. (1986), Specification Searches in Covariance Structure Modeling, *Psychological Bulletin* 100, 107-120.

Maccallum, R.C. (1995), Model Specifiaction, Procedures, Strategies and Related Issues, in Hoyle, R.H. (ed.), *Sructural Equation Modeling. Concepts, Issues and Applications*, Thousand Oaks, CA, Sage*,* (pp. 16-36).

Maccallum, R.C., Roznowski, M. and Necowitz, L.B. (1992), Model Modification in Covariance Structure Analysis, the Problem of Capitalization on Chance, *Psychological Bulletin* 111, 490-504.

Mahalanobis, P. (1936), On the Generalized Distance in Statistics, *Proceedings of the National Institute of Science, Calcutta, 12,* 49-55.

Martensen, A., Grønholdt, L. and Kristensen, K. (2000), The Drivers of Customer Satisfaction and Loyalty, Cross-Industry Findings From Denmark, *Total Quality Management* 11, 8544-8553.

Muthén, B. (1984), A General Structural Equation Model with Dichotomous, Ordered Categorical and Continuous Latent Variable Indicators. *Psychometrika* 43, 241-250.

Muthén, B. (1994), Multilevel Covariance Structure Analysis, *Sociological Methods and Research 22,* 376-398.

Muthén, B. (2001), Second-Generation Structural Equation Modeling with a Combination of Categorical and Continuous Latent Variables, New Opportunities for

Latent Class/Latent Growth Modeling, in Collins, L.M. and Sayer, A., (eds.), *New Methods for the Analysis of Change,* Washington, DC, APA, (pp. 291-322).

Muthén, B. and Kaplan, D. (1989), A Comparison of Some Methodologies for the Factor Analysis of Non-Normal Likert Variables. A Note on the Size of the Model, *British Journal of Mathematical and Statistical Psychology* 45, 19-30.

Muthén B., Kaplan, D. and Hollis. M. (1987), On Structural Equation Modeling with Data that Are not Missing Completely at Random, *Psychometrika* 52, 431-462.

Muthén, B. and Satorra, A. (1995), Complex Sample Data in Structural Equation Modeling, *Sociological Methodology* 25, 267-316.

Muthén, L. K. and Muthén, B. (2001), *Mplus User's Guide.* Los Angeles, CA, Muthén and Muthén.

Neale, M. C., Boker, S.M., Xie, G. and Maes, H. H. (1999), *Mx, Statistical Modeling. 5th Edition,* Box 126, MCV, Richmond, VA23298, Department of Psychiatry.

Nevitt, J. and Hancock, G. R. (2001), Performance of Bootstrapping Approaches to Model Test Ststaistics and Parameter Standard Error Estimation in Structural Equation Modelling, *Structural Equation Modeling* 8, 353-377.

Raykov, T. and Marcoulides, G.A. (2000), *A First Course in Structural Equation Modeling,* Mahwah, NJ, Lawrence Erlbaum.

Reichheld, F.F. and Sasser, W.E. (1990), Zero Defections, Quality Comes to Services, *Harvard Business Review* 68, 105-111.

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys,* New York, Wiley.

Satorra, A. (1990), Robustness Issues in Structural Equation Modeling, a Review of Recent Developments, *Quality and Quantity* 24, 367-86.

Satorra, A. (1992), Asymptotic Robust Inferences in the Analysis of Mean and Covariance Structures, in Marsden, P.V., (ed.), *Sociological Methodology 1992*, Cambridge, Basil Blackwell, (pp. 249-278).

Satorra, A. (1993), Asymptotic Robust Inferences in Multi-Sample Analysis of Augmented-Moment Matrices, in Rao, C.R. and Cuadras, C.M., (eds.), *Multivariate Analysis, Future Directions 2*, Amsterdam, North-Holland, (pp. 211-229).

Satorra, A. and Bentler, P.M. (1994), Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis, in Von Eye, A. and Clogg, C., (eds.), *Latent Variables Analysis, Applications to Developmental Research,* Thousand Oaks, Ca, Sage, (pp. 399-419).

Satorra, A. and Bentler, P.M. (1999), *A Scaled Difference Chi-Square Test Statistic for Moment Structure Analysis*, Technical Report. University of California, Los Angeles.

Du Toit, M., and du Toit, S., (2001), *Interactive LISREL, User's Guide,* Chicago, IL, Scientific Software International.

Widaman, K. (1993), Common Factor Analysis versus Principal Component Analysis, Differential Bias in Representing Model Parameters?, *Multivariate Behavioral Research* 25, 1-28.

Wold, H. (1975), Path Models with Latent Variables, the NIPALS Approach, in Blalock, H.M., Aganbegian, A., Borodkin, F.M., Boudon, R. and Cappecchi, V., (eds.), *Quantitative Sociology. International Perspectives on Mathematical and Statistical Modeling,* New York, Academic Press, (pp. 307-353).

Wothke, W. (2000), Longitudinal and Multi-Group Modeling with Missing Data. in Little, T.D., Schnabel, K.U. and Baumert, J., (eds.), *Modeling Longitudinal and Multilevel Data, Practical Issues, Applied Approaches and Specific Examples,* Mahwah, NJ, Lawrence Erlbaum, (pp. 219-240).

Yuan, K.-H., and Bentler, P.M. (2000), Three Likelihood-Based Methods for Mean and Covariance Structure Analysis with Nonnormal Missing Data, in Sobel, M.E. and Becker, M.P., (eds.), *Sociological Methodology 2000*, Washington, DC, American Sociological Association, (pp. 165-200).