

Factor Analysis and Discriminant Validity: A Brief Review of Some Practical Issues

Research Methods

Andrew M Farrell (corresponding and presenting author)

Aston Business School
Aston University
Birmingham
B4 7ET
UK
Tel: 00 44 (0) 121 204 4874
Fax: 00 44 (0) 121 204 4917
Email: a.m.farrell2@aston.ac.uk

John M Rudd

Aston Business School
Aston University
Birmingham
B4 7ET
UK
Tel: 00 44 (0) 121 204 3218
Fax: 00 44 (0) 121 204 4917
Email: j.m.rudd@aston.ac.uk

Abstract

Growth in availability and ability of modern statistical software has resulted in greater numbers of research techniques being applied across the marketing discipline. However, with such advances come concerns that techniques may be misinterpreted by researchers. This issue is critical since misinterpretation could cause erroneous findings. This paper investigates some assumptions regarding: 1) the assessment of discriminant validity; and 2) what confirmatory factor analysis accomplishes. Examples that address these points are presented, and some procedural remedies are suggested based upon the literature. This paper is, therefore, primarily concerned with the development of measurement theory and practice. If advances in theory development are not based upon sound methodological practice, we as researchers could be basing our work upon shaky foundations.

Keywords: Average Variance Extracted, Discriminant Validity, Factor Analysis, Latent Variable Modelling

Factor Analysis and Discriminant Validity: A Brief Review of Some Practical Issues

Introduction

It is natural to assume that all research contains some flaws; the myth of the perfect study represents something unattainable. However, without research, theoretical advances in the social sciences, and marketing in particular, would not occur. Consequently, the social science community needs to be confident that theoretical advances are arrived at through both sound conceptual argument and the application of rigorous and relevant methodological techniques. It is the application of two such methodological techniques that this paper focuses upon: factor analysis and discriminant validity testing.

Within the social sciences, latent variable modelling, and by implication structural equation modelling (SEM), has become an extremely popular modelling technique (Baumgartner and Homburg, 1996; Steenkamp and van Trijp, 1991). This is partially due to the efficacy of SEM techniques when following Churchill's (1979) scale development guidelines. The development or assessment of scales is often associated with the application of a factor analysis (either exploratory, EFA, or confirmatory, CFA) and further testing to establish the validity of measures (i.e., convergent, discriminant). In theory, other types of validity (e.g., face, nomological) should have been established prior to data collection. In that sense, we are concerned here with post-hoc measure development practice, and its implications for theory development and advancement.

Farrell (2009) calls for a review of discriminant validity assessment in organizational research. With this in mind, this paper has the two objectives of briefly reviewing: 1) the application of confirmatory factor analysis; and 2) the assessment of discriminant validity. Examples to illustrate points made in discussion are drawn from a brief review of the last two years of the *Journal of Marketing*, *Journal of Marketing Research*, and *Journal of Business Research*. In addition, other cases that the authors were aware of were also included. From this brief review of academic work, it appears that there are issues regarding the application of confirmatory factor analysis and discriminant validity assessment. The results of the review are presented in Table 1, and the examples will be elaborated upon shortly.

If a factor analysis is misinterpreted, and discriminant validity is not established, then measurement scales used in research may not function correctly, and conclusions made regarding relationships between constructs under investigation may be incorrect. For example, the strength of a relationship could be overestimated, or a relationship may be confirmed when in fact there is no real relationship (essentially, a Type II error). According to Armstrong (2009), researchers often fail to adequately read papers that they cite. Therefore, where theory development is concerned, there appears to be great faith that prior work has been appropriately conducted and reported although, as this paper shows, this is not always the case. Before we proceed with our review, there are certain concepts that require elaboration. These concepts are discriminant validity itself, shared variance, and the notion of average variance extracted (AVE).

Discriminant validity is the extent to which latent variable A discriminates from other latent variables (e.g., B, C, D). Discriminant validity means that a latent variable is able to account for more variance in the observed variables associated with it than a) measurement error or similar external, unmeasured influences; or b) other constructs within the conceptual

framework. If this is not the case, then the validity of the individual indicators and of the construct is questionable (Fornell and Larcker, 1981). Shared variance is the amount of variance that a variable (construct) is able to explain in another variable (construct). It is represented by the square of the correlation between any two variables (constructs). For example, if the correlation between two variables, x_1 and x_2 , is 0.6, then the shared variance between x_1 and x_2 is 0.36. If independent variables are correlated, they share some of their predictive power over dependent variables (Hair et al., 2006). The AVE estimate is the average amount of variation that a latent construct is able to explain in the observed variables to which it is theoretically related. A latent construct A will correlate with observed variables, x_1 and x_2 , that theoretically relate to A. This correlation is generally referred to as a factor loading. If we square each of these correlations, this gives the amount of variation in each observed variable that the latent construct accounts for (i.e., shared variance). When this variance is averaged across all observed variables that relate theoretically to a latent construct, we generate the AVE (Farrell, 2009).

There are similarities between AVE and shared variance. AVE is the average amount of variance in observed variables that a latent construct is able to explain, and shared variance is the amount of variance in observed variables *relating to another construct* that a latent construct is able to explain. Fornell and Larcker (1981) present a method for assessing the discriminant validity of two or more factors. Here, a researcher compares the AVE of each construct with the shared variance between constructs. If the AVE for each construct is greater than its shared variance with any other construct, discriminant validity is supported.

Objective 1: The Application of Confirmatory Factor Analysis (CFA)

When conducting a CFA, one should never be governed by the fit indices of the model alone. There are other factors to consider, such as the factor loading for each observed variable. Two brief examples follow. First, Tellis, Yin and Bell (2009, p. 11) present the results of a factor analysis in Table 4 of their work. This factor analysis was performed upon items that the authors generated from a review of the consumer innovativeness literature. From this literature review, the authors developed ten dimensions of innovativeness. They then adopted one item per dimension from the literature to measure each dimension. For example, for the novelty-seeking dimension, they reviewed novelty-seeking research and selected one item to act as an overall measure of that dimension. Hence, the items had been validated previously as part of separate scales, and this study therefore represented the first assessment of the revised innovativeness scale.

The fit indices for this factor analysis indicate an excellent fit (RMSEA = 0.04; CFI = 0.90; GFI = 0.99; AGFI = 0.98), and the authors present a three factor solution with the following factors: openness, enthusiasm, and reluctance. These factor names were assigned by the authors based upon the items which had loaded on each factor. But, although this factor analysis was conducted on an impressive sample of 5569 respondents from 15 countries, there are certain issues raised by the reporting. First, there is a negative factor loading for one item. Second, the factor loadings of many of the items are extremely low, ranging from -0.27 to 0.73, with a mean of 0.417. Recall that the square of a factor loading provides the amount of variance in the observed variable that the underlying construct is able to explain. Hence, low factor loadings (i.e., less than 0.7) result in situations where more than 50% of the variance in an observed variable is explained by factors other than the construct to which the variable is theoretically related (i.e., other constructs or types of error). As a result, such low factor loadings could be indicative of problems with the factor structure being represented.

However, because Tellis, Yin and Bell (2009) did not provide error terms for their factor loadings, we can not be certain if these issues are serious.

Because the information was not provided by Tellis, Yin and Bell (2009), we conducted a substitution exercise in order to provide expected estimates of the composite reliability (CR) and average variance extracted (AVE) results. Factor loadings (and their corresponding error terms) were taken from other, published, CFA results and substituted for the results presented in Tellis, Yin and Bell (2009). Composite reliability and AVE estimates were then calculated using the substituted data. Published CFA results were taken from Byrne (1998, p. 154), Diamantopoulos and Siguaw (2000, pp. 68-69), and Sharma (1996, p. 156). Where an equivalent factor loading was not present in the published CFA results, the next highest available loading was used. Using the next highest loading gives positively-biased CR and AVE estimates due to 1) higher factor loadings; and 2) lower error terms. Therefore, the results that follow provide a generous representation of possible CR and AVE estimates for the work of Tellis, Yin and Bell (2009). Results are presented in Table 2.

Table 2: CR and AVE Estimates for Tellis, Yin and Bell (2009) Based on Published CFAs

Source of Substituted Data	Openness		Enthusiasm		Reluctance	
	CR	AVE	CR	AVE	CR	AVE
Byrne (1998)	0.50	0.26	0.53	0.29	0.45	0.17
Diamantopoulos and Siguaw (2000)	0.63	0.36	0.68	0.42	0.68	0.35
Sharma (1996)	0.61	0.36	0.61	0.36	0.52	0.21

It should be noted that the substituted data that provided the closest approximation to the factor loadings presented by Tellis, Yin and Bell (2009) was that of Byrne (1998, p. 154). Such low CR and AVE estimates can be extremely problematic for interpretation of subsequent analysis (Farrell, 2009).

The second example is Tellis, Prabhu and Chandy's (2009) CFA of eight constructs, one of which is the dependent variable. The authors do not report CFA fit statistics, but state "the overall fit of the full model is satisfactory" (p. 12). In Appendix B of their paper, the authors list factor loadings, but not error terms (pp. 19-20). Factor loadings across the eight constructs range from 0.32 to 0.88. Substituting data from Byrne (1998, p. 154), which again provided the closest match to the published factor loadings, expected AVEs for the eight constructs are: 0.27, 0.34, 0.34, 0.35, 0.42, 0.58, 0.58 and 0.68. Notably, the dependent variable has an AVE of 0.42, and five of the remaining seven constructs have significant effects on this variable. If factor loadings in a CFA are low, as in the cases presented here, researchers should conduct EFA and examine item cross-loadings. Cross-loading items represent prime candidates for removal from subsequent analysis with the goal of improving model fit.

An additional concern is that it is not uncommon to see authors report that they assessed convergent and discriminant validity through confirmatory factor analysis (Luo, Kannan and Ratchford, 2008; Noriega and Blair, 2008; Orth and Malkewitz, 2008; Voss and Voss, 2008). While it is true that CFA does assess convergent validity, it is not the best technique to assess discriminant validity. However, earlier work in the area of structural equation modeling has indicated that CFA does assess discriminant validity (Bagozzi, Yi and Phillips, 1991). The

important thing to bear in mind here is that each structural equation model represents only one possible fit to the data, so unless authors are testing competing models (with each model supported by sound theory), they can not be certain that their model provides the only fit to the data. CFA does not provide evidence of cross-loading items (a major source of insufficient discriminant validity) and also does not provide direct evidence of AVE estimates. Therefore, instead of only conducting a CFA, researchers should first conduct EFA to identify cross-loading items (i.e., for subsequent removal from the analysis if necessary). Hence, CFA should be used to confirm factor structure, while EFA should be used to identify potentially problematic items that might subsequently cause poor CFA fit. Researchers should then calculate AVE and shared variance estimates for each construct. AVE and shared variance estimates allow for the performance of the Fornell and Larcker (1981) discriminant validity test, a more stringent test than the often-used paired construct test (Farrell, 2009). To summarise this section: researchers should not only evaluate a CFA based upon model fit statistics but should also pay close attention to factor loadings, and CFA should not be used standalone to assess convergent and discriminant validity, as it is not the most stringent test for discriminant validity.

Objective 2: The Assessment of Discriminant Validity

Whenever there are high construct inter-correlations, there is a need to assess discriminant validity, in order to have confidence in subsequent research findings (Farrell, 2009). In Table 1, there are three examples given where authors report high construct inter-correlations but do not attempt to assess discriminant validity (Chitturi, Raghunathan and Mahajan, 2008; Morgan and Rego, 2009; Srinivasan, Pauwels, Silva-Risso and Hanssens, 2009). Ideally, authors faced with high construct inter-correlations would look to conduct some form of discriminant validity assessment on the constructs involved, to give greater confidence to later interpretation of findings.

There are a number of ways to assess discriminant validity between constructs. For example, researchers can conduct a paired construct test (Jorsekog, 1971), apply the Fornell and Larcker (1981) technique, or conduct a multi-trait multi-method evaluation of constructs. However, given limitations in data collection, and a need for more stringent evaluations of validity, it appears that the Fornell and Larcker (1981) technique represents the best method to apply (Farrell, 2009). Using this technique, for discriminant validity to be supported Hair et al. (2006, p. 778) note that "the variance extracted estimates should be greater than the squared correlation estimate" and Fornell and Larcker (1981, pp. 45-46) indicate that for any two constructs, A and B, the AVE for A and the AVE for B both need to be larger than the shared variance (i.e., square of the correlation) between A and B. That is, both AVE estimates have to be greater than the shared variance estimate. Despite these recommendations, it is not uncommon to find discriminant validity assessment arguments such as that below:

"Discriminant validity is assessed by comparing the shared variance (squared correlation) between each pair of constructs against the average of the AVEs for these two constructs"

(Bove et al., 2009, p. 702; Hassan et al., 2007; Walsh, Beatty and Shiu, 2009)

Clearly, when comparing this argument to the passages drawn from Hair et al. (2006) and Fornell and Larcker (1981), there is a misinterpretation of what is actually being advocated. In the case of Bove et al. (2009), a lack of discriminant validity actually causes significant problems for the remainder of the analysis (see Farrell, 2009, for further details). Styles,

Patterson, and Ahmed (2008) also report findings that suffer from a lack of discriminant validity, leading to problems in the interpretation of three out of the 21 hypotheses tested.

Brasel and Gips (2008) used four multi-item scales to assess affective response to advertising, brand memory, stopping strategy, and opinions of advertising overall. Similarly, Noriega and Blair (2008) used four multi-item scales to measure attitude toward the ad, attitude toward the brand, purchase intentions, and involvement with the product class. However, Brasel and Gips (2008) did not perform any form of convergent or discriminant validity analysis on the four scales. Noriega and Blair (2008) did perform a CFA on their measurement scales, but they provide no detail of AVE or shared variance assessment between the four scales. Instead, they appear to assume that the CFA alone indicates that their four scales perform well, which is not necessarily the case (see Objective 1 above).

When assessing the use of measurement scales in marketing research, authors should pay close attention to construct inter-correlations, which provide an indication of whether to expect discriminant validity problems. Researchers should then conduct EFA, to identify cross-loading items. Such items are prime candidates for subsequent discriminant validity issues. Once EFA has been conducted, researchers should perform CFA, paying close attention to factor loadings (see Objective 1), before calculating AVE and shared variance estimates. AVE and shared variance estimates should be compared to assess discriminant validity (Fornell and Larcker, 1981). Often in journal articles, because none or only some of this information is reported, we do not know whether constructs adequately discriminate from each other. As such, there could be interpretation issues in the subsequent analysis.

Conclusions and Recommendations

This paper set out to highlight potential issues regarding: 1) the use of confirmatory factor analysis in marketing research; and 2) the assessment of discriminant validity in marketing research. Examples have been provided that illustrate situations where CFA has been misinterpreted, and discriminant validity has either not been adequately assessed or has not been assessed at all. Should the responsibility for correct procedure lie with researchers or with journal review boards, who can request further analysis or presentation of certain material? The answer to this question is outside of the bounds of this paper, but is perhaps one for future researchers to consider. The ramifications for misinterpretation of research techniques could be serious, given our reliance upon the work of others. If we are citing previous work, perhaps using it as the conceptual foundation of our own, then we need to be sure that the analysis reported in that previous work has been correctly reported, otherwise we run the risk of propagating misinterpretation. If advances in theory development are not based upon sound methodological practice, we as researchers could be basing our work upon shaky foundations. In this respect, we believe in the correctness of holding all research to rigorous standards (Reibstein, Day and Wind, 2009). However, that is not to say that we believe that rigorous methodological standards should take the place of sound conceptual underpinnings. Rather, we believe that all research can benefit from sound application in both areas.

This comment sought to draw attention to problems related to the misinterpretation of two commonly applied research techniques: confirmatory factor analysis and discriminant validity assessment. It presented examples of misinterpretation of factor analysis and confusion surrounding the Fornell and Larcker (1981) discriminant validity test. Ideally, researchers will take more care when applying and interpreting psychometric assessment in future, essentially

extending their work past the standard two-step model evaluation process advocated by Anderson and Gerbing (1988). It is hoped that this article will serve to encourage more conscientious application and interpretation of factor analysis and discriminant validity testing. This should enable researchers to use past work with greater confidence, resulting in stronger theory development in the future.

References

Anderson, J.C., Gerbing, D.W., 1988. Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin* 103 (3), 411-423.

Armstrong, J.S., 2009. Summary of major research findings. Located at: <http://marketing.wharton.upenn.edu/documents/misc/Major%20Findings.doc> (Accessed 22 June 2009: 17:42)

Bagozzi, R.P., Yi, Y., Phillips, L.W., 1991. Assessing construct validity in organizational research. *Administrative Science Quarterly* 36, 421-458.

Baumgartner, H., Homburg, C., 1996. Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing* 13, 139-161.

Bove, L.L., Pervan, S.J., Beatty, S.E., Shiu, E., 2009. Service worker role in encouraging customer organizational citizenship behaviors. *Journal of Business Research* 62 (7), 698-705.

Brasel, S.A., Gips, J., 2008. Breaking through fast-forwarding: Brand information and visual attention. *Journal of Marketing* 72 (November), 31-48.

Byrne, B.M., 1998. *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*. Lawrence Erlbaum Associates, Mahwah, NJ.

Chitturi, R., Ranhunathan, R., Mahajan, V., 2008. Delight by design: The role of hedonic versus utilitarian benefits. *Journal of Marketing* 72 (May), 48-63.

Churchill, G.A., 1979. A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research* 16 (February), 64-73.

Diamantopoulos, A., Siguaw, J.A., 2000. *Introducing LISREL*. Sage Publications, London.

Farrell, A.M., 2009. Insufficient discriminant validity: A comment on Bove, Pervan, Beatty and Shiu (2009). *Journal of Business Research*, in press.

Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18 (1), 39-50.

Hair, Jr., J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., 2006. *Multivariate data analysis* (6th Ed.), Pearson-Prentice Hall, Upper Saddle River, NJ.

- Hassan, L.M., Walsh, G., Shiu, E.M.K., Hastings, G., Harris, F., 2007. Modeling persuasion in social advertising. *Journal of Advertising* 36 (2), 15-31.
- Joreskog, K.G., 1971. Statistical analysis of sets of congeneric tests. *Psychometrika* 36 (2), 109-133.
- Luo, L., Kannan, P.K., Ratchford, B.T., 2008. Incorporating subjective characteristics in product design and evaluations. *Journal of Marketing Research* 45 (April), 182-194.
- Morgan, N.A., Rego, L.L., 2009. Brand portfolio strategy and firm performance. *Journal of Marketing* 73 (January), 59-74.
- Noriega, J., Blair, E., 2008. Advertising to bilinguals: Does the language of advertising influence the nature of thoughts? *Journal of Marketing* 72 (September), 69-83.
- Orth, U.R., Malkewitz, K., 2008. Holistic package design and consumer brand impressions. *Journal of Marketing* 72 (May), 64-81.
- Reibstein, D.J., Day, G.S., Wind, J., 2009. Guest Editorial: Is marketing academia losing its way? *Journal of Marketing* 73 (4), 1-3.
- Sharma, S., 1996. *Applied Multivariate Techniques*. John Wiley & Sons, Inc., New York.
- Srinivasan, S., Pauwels, K., Silva-Risso, J., Hanssens, D.M., 2009. Product innovations, advertising, and stock returns. *Journal of Marketing* 73 (January), 24-43.
- Steenkamp, J-B.E.M., van Trijp, H.C.M., 1991. The use of LISREL in validating marketing constructs. *International Journal of Research in Marketing* 8, 283-299.
- Styles, C., Patterson, P.G., Ahmed, F. 2008. A relational model of export performance. *Journal of International Business Studies* 39, 880-900.
- Tellis, G.J., Prabhu, J.C., Chandy, R.K., 2009. Radical innovation across nations: The preeminence of corporate culture. *Journal of Marketing* 73 (1), 3-23.
- Tellis, G.J., Yin, E., Bell, S., 2009. Global consumer innovativeness: Cross-country differences and demographic commonalities. *Journal of International Marketing* 17 (2), 1-22.
- Voss, G.B., Voss, Z.G., 2008. Competitive density and the customer acquisition-retention trade-off. *Journal of Marketing* 72 (November), 3-18.
- Walsh, G., Beatty, S.E., Shiu, E.M.K., 2009. The customer-based corporate reputation scale: Replication and short form. *Journal of Business Research*, in press.

Table 1: Issues in Discriminant Validity and Confirmatory Factor Analysis Assessment

Authors	Deviations from Suggested Practice	Suggested Procedure *
Tellis, Yin and Bell (2009)	Low factor loadings in CFA, negative factor loading in CFA, probable low reliabilities and AVEs	Conduct EFA, examine modification indices and item cross-loadings, use CFA outputs to calculate AVE measures
Tellis, Prabhu and Chandy (2009)	Low factor loadings in CFA, probable low reliabilities and AVEs	Conduct EFA, examine modification indices and item cross-loadings, use CFA outputs to calculate AVE measures
Bove, Pervan, Beatty and Shiu (2009)	Insufficient discriminant validity displayed. Misinterpretation of Fornell and Larcker (1981)	Conduct an EFA and examine item cross-loadings
Styles, Patterson and Ahmed (2008)		
Brasel and Gips (2008)	Employ four multi-item scales, but do not conduct EFA, CFA, or discriminant validity assessment	Conduct EFA and CFA, use CFA outputs to calculate AVE, compare AVE to shared variance estimates
Luo, Kannan and Ratchford (2008)	Employ four multi-item scales, but do not conduct discriminant validity assessment, assume CFA measures discriminant validity	Use CFA outputs to calculate AVE, and compare AVE to shared variance estimates
Noriega and Blair (2008)		
Orth and Malkewitz (2008)		
Voss and Voss (2008)		
Chitturi, Raghunathan and Mahajan (2008)	Report high construct inter-correlations (0.78, 0.81, 0.84, 0.86) and do not conduct discriminant validity tests	Attempt to assess the discriminant validity of the highly inter-correlated constructs
Morgan and Rego (2009)	Report high construct inter-correlation (0.856) and do not conduct discriminant validity tests	
Srinivasan, Pauwels, Silva-Risso and Hanssens (2009)	Report high construct inter-correlation (0.92) and do not conduct discriminant validity tests	

* Suggested Procedures are drawn from the work of Anderson and Gerbing (1988); Farrell (2009); Fornell and Larcker (1981); and Jorsekog (1971)