Insufficient discriminant validity:

A comment on Bove, Pervan, Beatty, and Shiu (2009)

Dr Andrew M. Farrell, Loughborough University

December 2008

The author expresses his gratitude to John Cadogan; his helpful and insightful comments on previous drafts of this manuscript were invaluable. Send correspondence to Andrew M. Farrell, Marketing and Retailing Research Group, Business School, Loughborough University, Ashby Road, Loughborough, Leics LE11 3TU, UK, Tel: 00 44 (0) 1509 22 31 76, Fax: 00 44 (0) 1509 22 27 23 (Email: a.m.farrell@lboro.ac.uk).

Insufficient discriminant validity:

A comment on Bove, Pervan, Beatty, and Shiu (2009)

Abstract

Bove, Pervan, Beatty, and Shiu (2009) develop and test a latent variable model of the role of service workers in encouraging customers' organizational citizenship behaviors. However, Bove et al. (2009) claim support for hypothesized relationships between constructs that, due to insufficient discriminant validity regarding certain constructs, may be inaccurate. This research comment discusses what discriminant validity represents, procedures for establishing discriminant validity, and presents an example of inaccurate discriminant validity assessment based upon the work of Bove et al. (2009). Solutions to discriminant validity problems and a five-step procedure for assessing discriminant validity then conclude the paper. This comment hopes to motivate a review of discriminant validity issues and offers assistance to future researchers conducting latent variable analysis.

Keywords: average variance extracted, discriminant validity, latent variable modeling

1. Introduction

Discriminant validity establishment is crucial for conducting latent variable analysis (Bollen, 1989; Fornell and Larcker, 1981). Without it, researchers can not be certain whether results confirming hypothesized structural paths are real or whether they are a result of statistical discrepancies. However, the establishment of discriminant validity is not always present in articles across the marketing literature. For example, Bove et al. (2009) incorrectly apply the Fornell and Larcker (1981) discriminant validity test and, as a result, the authors might have drawn incorrect conclusions regarding relationships in their structural model.

This present comment sheds light upon the assessment of discriminant validity issues. The next section briefly introduces the concepts of discriminant validity, shared variance and average variance extracted. The following section provides details of how to assess discriminant validity, followed by problems that can arise if discriminant validity is not confirmed, using examples from Bove et al. (2009). The paper then concludes with a five-step procedure for addressing discriminant validity issues.

2. Discriminant Validity, Shared Variance, and 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, Black, Babin, Anderson, and Tatham, 2006). Inspection of the correlation matrix between latent constructs can often identify potential shared variance issues.

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.

For example, latent variable A correlates with two observed variables, x_1 and x_2, at the level of 0.9 (i.e., both variables load on A at a value of 0.9). Squaring the correlations, we generate the shared variance between A and x_1 and x_2, 0.81 in both cases. In this example, the average variance extracted by A in x_1 and x_2 would therefore be 0.81 (notwithstanding measurement error, discussed later). 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.

3. Establishing Discriminant Validity

3.1 Paired Constructs Test

Anderson and Gerbing (1988) suggest that the parameter estimate for two factors be constrained to 1.0 (constrained model) and compared to a model where this parameter is freely estimated (unconstrained model). This test is then run for every possible pairing of constructs in a study. If the unconstrained model, with the drop of one degree of freedom, returns a chi-square value that is at least 3.84 lower than the constrained model, then a two factor solution provides a better fit to the data, and discriminant validity between A and B is supported.

3.2 Average Variance Extracted versus Shared Variance Test

Fornell and Larcker (1981) also 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.

Bove et al. (2009) noted that “discriminant validity is assessed by comparing the shared variance (squared correlation) between each pair of constructs against the average of the AVEs for these constructs.” However, 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, not the average of the AVE estimates, as argued by Bove et al. (2009). Had the Fornell and Larcker (1981) technique been applied correctly in Bove et al. (2009) it would have demonstrated insufficient discriminant validity between certain constructs, and the structural model should not have been assessed until this issue was addressed (see Section 4).

Particularly relevant to the shared variance versus AVE technique is the notion of measurement error. One of the benefits of using structural equation modeling (SEM) is that it enables a researcher to account for measurement error in variables (Bollen, 1989). An important point to note is that when measurement error is taken into account, correlations between variables generally, though not always, increase in magnitude (Grewal, Cote, and Baumgartner, 2004). The formula for the calculation of AVE contained in Fornell and Larcker (1981, p. 46) requires measurement error terms from the CFA output. For the purposes of assessing discriminant validity, it is better to use the CFA correlation matrix, as to use a correlation matrix without measurement error taken into account (i.e., from a program such as SPSS or PRELIS) could lead to misleading results. If an AVE that includes measurement error is compared to a shared variance estimate that does not, then the AVE is being compared to a potentially downwardly-biased shared variance estimate, and the test may erroneously conclude that the variables discriminate. Use of AVE and shared variance estimates that account for measurement error therefore provides a more stringent evaluation of the AVE versus squared correlation test.

3.3 Multitrait-Multimethod Matrix (MTMM)

The MTMM method uses more than one measure of constructs (i.e., multitrait) and more than one method to measure them (i.e., multimethod) in order to assess both convergent and discriminant validity (Bollen, 1989). By collecting data on constructs using at least two separate traits and methods, it is easier to identify discriminant validity problems. The major drawbacks of this method for researchers are that it is cumbersome, requires more data collection, and may suffer from interpretation issues (Bollen, 1989).

4. Insufficient Discriminant Validity: An Example using Bove et al. (2009)

If discriminant validity is not established, then 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. Bove et al. (2009) generate AVE estimates and shared variance estimates for five constructs: commitment to the service worker, credibility of the service worker, benevolence of the service worker, personal loyalty, and customer organizational citizenship behaviors (OCBs). Table 1 is developed using Bove et al.'s (2009) Tables 2 and 3.

Table 1 here.

Table 1 identifies three cases where there is insufficient discriminant validity (a, b, c). In case a (Fig. 1), it can be seen that the AVE of 0.87 for commitment (ξ_1) is greater than the shared variance of 0.67 (i.e., 0.82^2) between commitment and personal loyalty (ξ_2). However, the AVE for personal loyalty (ξ_2) is only 0.65, lower than 0.67, the shared variance between commitment and personal loyalty. Hence, the commitment latent variable (ξ_1) explains more of the variance in observed variables x_3, x_4 and x_5 than does the personal loyalty latent variable (ξ_2), despite the fact that x_3, x_4 and x_5 are supposed to be measures of personal loyalty (not commitment). As a result, one is not sure whether x_3, x_4 and x_5 are really very good measures of personal loyalty. Of course, this reduces confidence in the authors' model containing personal loyalty and commitment. A similar situation exists for the other two cases, b and c.

Figure 1 here.

If discriminant validity is not established, then latent constructs are having an influence on the variation of more than just the observed variables to which they are theoretically related. Of course, if one latent construct is measuring two different concepts, A and B, it is tautological that A could have a positive relationship with B, or vice versa. When the three cases identified above are considered as part of Bove et al.'s (2009, Fig. 1) path diagram, discriminant validity issues result in problems with interpretation of the findings relating to these latent constructs. More specifically, hypotheses H1, H4 and H5a are problematic. In addition, because of discriminant validity issues, a researcher can not be certain exactly which latent construct credibility of the service worker and benevolence of the service worker are acting as antecedent to. Hence, there are interpretation issues regarding hypotheses H3, H5b and H5c. This is unfortunate, as Bove et al.'s (2009) work makes a strong theoretical contribution to an especially interesting field of research; however, due to insufficient discriminant validity, the analysis within the paper weakens the results.

5. Suggestions for Lack of Discriminant Validity

Introduction of a common method factor may help to reduce variance inflation, reducing shared variance estimates between latent constructs and observed variables. Consequently, the likelihood that AVE estimates will be greater than shared variance estimates increases. However, it increases model complexity. One can also conduct further analysis using residual terms, partialling out shared variance (e.g., Little, Bovaird, and Widaman, 2006), or use other techniques such as tolerance analysis (Nunnally and Bernstein, 1994).

EFA is useful for learning if discriminant validity issues are a result of poorly performing (i.e., cross-loading) items. If items cross-load on more than one latent variable, removal of offending items should improve discriminant validity. CFA can also be used, and a researcher can inspect modification indices or correlated error terms, but the ability of EFA to identify cross-loadings is particularly beneficial. During item removal tests, a researcher needs to be aware of the trade-off between the number of scale items (for face validity or construct coverage) or measurement scales that perform well and discriminate.

If discriminant validity issues persist, no option may exist but to combine constructs into one overall measure, as in the case of transformational leadership, where correlations in the region of 0.8 to 0.9 are regularly reported in the literature between dimensions that are theoretically distinct. In such cases, researchers tend to collapse transformational leadership measures into a single construct, rather than conduct dimension-by-dimension analysis. Depending on the nature of the constructs under investigation, this technique may not always be appropriate. In Bove et al. (2009), it does not make theoretical sense to combine commitment to the service worker, personal loyalty, and/or customer OCBs into one overall measure.

If none of the methods presented address the issue, a researcher may have to collect additional data to determine if discriminant validity or multicollinearity issues are a result of sampling flukes (Bollen, 1989). If problems still persist, Cohen, Cohen, West, and Aiken (2003) suggest dropping one (or more) independent variables (i.e., collinear variables that demonstrate insufficient discriminant validity) from the regression equation(s).

6. Conclusions and Recommendations

Implications of a lack of discriminant validity have been presented, demonstrating that a lack of discriminant validity reduces confidence in results, and it is hoped that researchers will place more emphasis upon assessing discriminant validity. A five-point plan follows for the assessment of discriminant validity.

(1) Researchers should perform EFA to identify items which cross-load. Researchers can also perform CFA and examine modification indices or correlated error terms, which again serve to identify cross-loading items. Examination of the standardized PHI matrix is useful here.

(2) As part of CFA, researchers should perform paired correlation tests.

(3) Researchers should calculate AVE estimates for each construct under investigation.

(4) Researchers should perform the AVE versus shared variance test (Fornell and Larcker, 1981), making sure that shared variance estimates are taken from structural equation output so as to include measurement error.

(5) If discriminant validity is not sufficiently established as a result of steps one through four, researchers should implement the suggestions in Section 5 before proceeding with structural model analysis.

Of course, presentation of information relating to validity testing can increase article length. There are two points to be made on this issue. Firstly, the reporting of discriminant validity statistics in the form of correlation and shared variance matrices can serve as an aid to meta-analysts when they seek to gather information from published work. Secondly, it may be that journals in future look to supplement their hardcopy issues with online content (e.g., *Marketing Science*). Such online content could reduce page space, and serve as appendices for drawn out statistical procedures, and preserve the information-rich content of an article.

This comment seeks to draw attention to problems of discriminant validity in latent variable modeling research. It has highlighted the concepts of average variance extracted and shared variance, and has documented procedures to assess the discriminant validity of latent variables prior to structural analysis. Consequences of insufficient discriminant validity have been shown using a real-life example and procedures to address these problems have been presented. It is hoped that this article will serve to encourage assessment of discriminant validity by researchers with the aim of increasing the overall psychometric performance of latent variable models in the future.

References

Anderson, JC, Gerbing, DW. Structural equation modeling in practice: A review and recommended two-step approach. Psych Bull 1988;103(3):411-23.

Bollen, KA. Structural equation models with latent variables. New York: John Wiley & Sons; 1989.

Bove, LL, Pervan, SJ, Beatty, SE, Shiu, E. Service worker role in encouraging customer organizational citizenship behaviors. J Bus Res 2009;62(7):698-705.

Cohen, J, Cohen, P, West, SG, Aiken, LS. Applied multiple regression / correlation for the behavioral sciences. Mahwah, NJ: Lawrence Erlbaum Associates; 2003.

Fornell, C, Larcker, DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981;18(1):39-50.

Grewal, R, Cote, JA, Baumgartner, H. Multicollinearity and measurement error in structural equation models: Implications for theory testing. Mark Sci 2004;23(4):519-529.

Hair, Jr., JF, Black, WC, Babin, BJ, Anderson, RE, Tatham, RL. Multivariate data analysis (6th Ed.) Upper Saddle River, NJ: Pearson-Prentice Hall; 2006.

Little, TD, Bovaird, JA, Widaman, KF. On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. Struct Equ Model 2006;13(4):497-519.

Nunnally, JC, Bernstein, IH. Psychometric theory (3rd Ed.) New York: McGraw-Hill. Inc. 1994.

Podsakoff, PM, MacKenzie, SB, Lee J-Y, Podsakoff, NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. J App Psychol 2003;88(5):879-903.

Table 1

Average Variance Extracted and Shared Variance Estimates

	Variable	Items	1	2	3	4	5
1	Commitment to Service Worker	2	**0.87**	0.26	0.41	0.67 (a)	0.58 (b)
2	Credibility of Service Worker	4	0.51	**0.66**	0.59	0.52	0.29
3	Benevolence of Service Worker	3	0.64	0.77	**0.63**	0.62	0.45
4	Personal Loyalty	3	0.82	0.72	0.79	**0.65**	0.61 (c)
5	Customer OCBs	8	0.76	0.54	0.67	0.78	**0.55**

NOTE: Correlations are below the diagonal, squared correlations are above the diagonal, AVE estimates are presented on the diagonal.

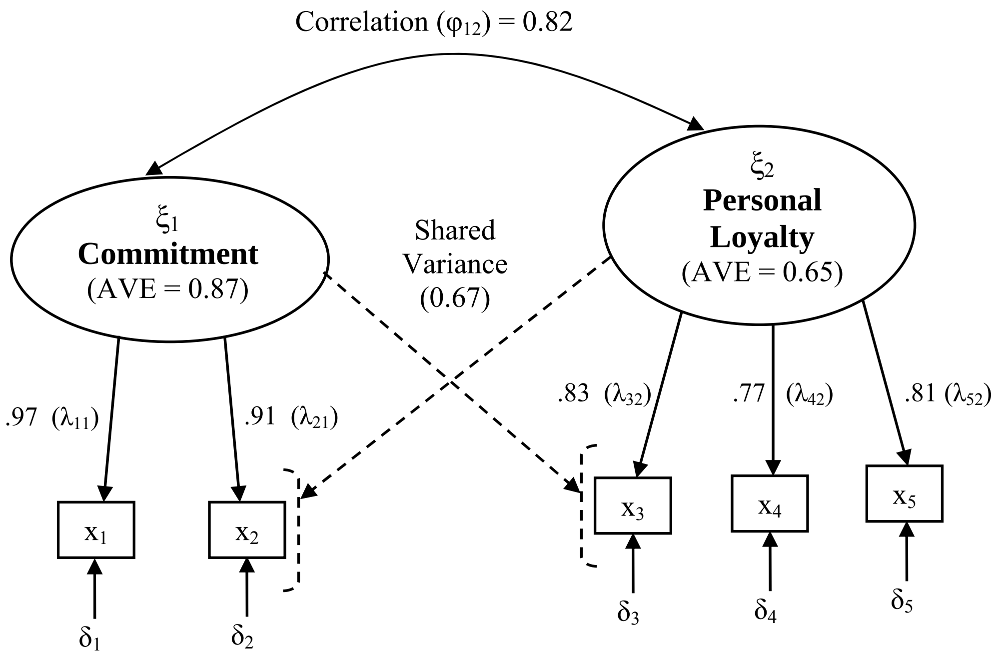

Fig. 1. Example of average variance extracted versus shared variance test.