JITTA

JOURNAL OF INFORMATION TECHNOLOGY THEORY AND APPLICATION

ISSN: 1532-3416

Understanding the Role of Theory on Instrument Development: An Examination of Strengths and Weaknesses of Discriminant Validity Analysis Techniques

Damon E. Campbell
Else School of Management
Millsaps College
damon.campbell@millsaps.edu

Damon E. Campbell
Else School of Management
Millsaps College

John D. Wells
Isenberg School of Management
University of Massachusetts, Amherst

Craig Parks
Department of Psychology
Washington State University

Abstract:

Numerous researchers have called attention to many important issues in instrument development throughout the relatively short history of the information systems (IS) academic research discipline (e.g., Petter, Straub, & Rai 2007; Straub, Boudreau, & Gefen 2004; Straub 1989). With the accumulation of knowledge related to the process of instrument development, it has now become necessary to take a closer look at specific aspects of this process. This paper focuses on construct validity, specifically discriminant validity, and examines some popular methods of supporting this type of validity when using cross-sectional data. We examine strengths and weaknesses of these analysis techniques with a focus on the role of theory and informed interpretation. We highlight the applicability of these techniques by analyzing a sample dataset where we theorize two constructs to be highly correlated. With this paper, we provide both researchers and reviewers a greater understanding of the highlighted discriminant validity analysis techniques.

Keywords: Construct Validity, Quantitative Analysis, Discriminant Validity, Average Variance Extracted, χ^2 Analysis.

Jerry Chang was the Senior Editor for this paper.

1 Introduction

Throughout the years, the information systems (IS) discipline has raised issues regarding survey instrument development (e.g., Gefen, Straub, & Boudreau, 2000; Lewis, Templeton, & Byrd, 2005; MacKenzie, Podsakoff, & Podsakoff, 2011; Petter, Straub, & Rai, 2007; Straub, Boudreau, & Gefen, 2004; Straub, 1989). IS research commonly uses cross-sectional data; as such, numerous techniques have been developed to validate instruments. Previous culminating works in this line of research include Straub et al.'s (2004) and Lewis et al.'s (2005) papers that propose validation guidelines for measurement instruments. A paper in a recent special issue of *MIS Quarterly* has also stressed the need to integrate and disseminate advancements in these areas to accumulate knowledge (MacKenzie et al., 2011). Researchers have also focused on specific issues regarding instrument development and validation (e.g., formative vs. reflective construct measurement, and common method bias) (Bagozzi, 2011; Bollen 2011; Diamantopoulos 2011; Petter et al., 2007; Straub & Burton-Jones, 2007). Following this stream of research, we focus on the area of discriminant validity analysis—specifically on the role of theory and the strengths and weaknesses of some common analysis techniques.

We examine some analysis techniques that are confirmatory (i.e., correlation-based, average variance extracted, χ^2 difference tests) as opposed to content-driven or item-selection methods along with their inherent strengths and weaknesses. With this understanding, we provide researchers the tools to use these techniques faithfully and interpret their results appropriately. We do not focus on invalidating any of the analysis techniques but compare and contrast different types of discriminant validity analysis. In doing so, we highlight the importance of theoretically informed interpretations of results and minimize the strict adherence to blind statistical procedure and subjective rules-of-thumb.

To accomplish these goals, we first discuss the role of discriminant validity analysis under the larger context of instrument development in Section 2. In Section 2.1, we discuss some of the difficulties in establishing discriminant validity. In Sections 3 and 4, we use a sample dataset to illustrate a process of instrument development. The research study used in this example is theoretically justified and includes cross-sectional data containing two highly correlated constructs. Although the situation is not typical, the dataset does provide key illustrations of how IS researchers can effectively use construct validation procedures and guidelines. During this process, the question of discriminant validity arises and we feature the strengths and weaknesses of some discriminant validity analysis techniques. In Section 5, we outline a method for assessing discriminant validity and summarize the contributions and limitations of our research.

2 Developing Survey Items

A fundamental assumption of instrument development is that, depending on its orientation, research can be theory driven (i.e., confirmatory) or data driven (i.e., exploratory). We focus on a theory-driven approach to instrument development. Although data-driven techniques can be used in this process, the distinction between exploratory and confirmatory approaches is important (Hughes, Price, & Marrs, 1986). Choosing between these approaches is often necessary when building theory versus testing theory. Given a strict adherence to the theory-testing paradigm, this paper emphasizes the critical role of a strong conceptual understanding in developing survey items.

Contribution:

This work makes two primary contributions. First, it descriptively outlines the role theory has in instrument development and the commonly used methods in IS literature for establishing discriminant validity. This paper compares and contrasts these analysis techniques to inform researchers of the strengths and weaknesses of each. Such an understanding is essential for researchers in interpreting the results of data analysis. Second, this paper contributes to a prescriptive method of analyzing discriminant validity in situations of highly correlated and theoretically justified relationships between constructs. This method, summarized in Figure 5, prescribes how analysis techniques can be used to maximize their strengths. This research supports the need for researchers to avoid an inexperienced reliance on procedural aspects of any given discriminant validity analysis. We also stress the importance and use of theory in interpreting what the data reveals. We do not propose that researchers go beyond their data but instead use theory and triangulation of multiple analysis techniques in analyzing data.

Note that the data analyzed in social science research are often observable reactions (i.e., responses) to stimuli (i.e., indicators or manifest variables, such as items on a questionnaire) that are designed to capture and represent a true score of a construct (i.e., latent variable) that is both unobservable and impossible to directly measure. The true score, as defined for our purposes here, refers to the actual level of a latent variable that, by definition, cannot be directly observed or measured. One draws appropriate indicators from the theory being tested. In turn, the first step in the instrument-development process is defining the construct(s). Effective construct development constitutes a theoretical exercise requiring an *a priori* conceptual understanding of the construct and its relationship with other constructs. A fundamental lack of theoretical support when identifying construct indicators (i.e., items) results in poorly constructed measures.

Using scores and indicators necessitates that the researcher justify the theoretical inferences drawn (Messick, 1981). Such a justification has two prerequisites: (1) reliability and (2) validity. To provide a framework for discussing statistical analysis and interpretation, we review the common practices upheld in the literature about developing and assessing indicators (see Campbell & Fiske, 1959; Gefen et al., 2000; Lewis et al. 2005; Petter et al., 2007; Straub et al., 2004; Straub, 1989). Initially, one assesses indicators in comparison to the conceptual definition of the construct for content validity. Content validity refers to whether the indicators fall in the conceptual domain of the latent variable (Shadish, Cook, & Campbell, 2002; Straub et al., 2004). Once items are used in a survey, one conducts other assessments of reliability and validity.

Reliability refers to the rate to which an item garners a consistent response. Such an assessment ensures that items are internally consistent and consistent across time and other contexts. Such reliability reflects a construct's theoretical significance when respondents consistently interpret individual items. For example, researchers use test-retest reliability methods to assess consistency of responses over time, and they often use generalizability studies to test responses across contexts. Another related aspect to reliability is unidimensionality. Unidimensionality refers to "the existence of one latent trait or construct underlying a set of measures" (Anderson, Gerbing, & Hunter, 1987, p. 432). For example, a strawberry has many qualities (e.g., color, size, and flavor). If one developed a measurement instrument to gauge perceptions of strawberries that contained items that focused in one of these different qualities rather than a general "strawberry-ness", results may indicate more than a single dimension to the instrument.

Another area of utmost concern for instrument development is in assessing validity. A reliable yet invalid or valid yet unreliable measure is useless. Validity refers to "the approximate truth of an inference" or "a judgment about the extent to which relevant evidence supports that inference as being true or correct" (Shadish et al., 2002, p. 34). Although there are other areas of validity[1], here we focus on those most relevant to instrument development. As we have already defined content validity, we now elaborate on the role of convergent and discriminant validity in terms of construct validity.

Construct validity is the broadest of these concepts in that it subsumes the others. Nunnally et al. (1994, p. 86) state that construct validity consists of: "1) specifying the domain of observables related to the construct [content validity]; 2) determining the extent to which observables tend to measure the same thing [convergent validity], several different things, or many different things [discriminant validity]". Convergent validity is the extent to which a latent variable's observed variables correlate highly with other observed variables that represent the same latent variable (Straub et al., 2004). Convergent validity analysis techniques provide statistical evidence that key construct properties are considered to be theoretically congruent. Statistical assessments of reliability, content validity, and convergent validity are normally focused at the indicator level. However, assessments of discriminant validity are often focused at the construct level. Following our previous example related to strawberry-ness, convergent validity would assess an items' ability to be like other strawberry items, while discriminant validity would be assessing a measure to ensure it is unlike other measures for different (i.e., not strawberry) constructs.

"Discriminant validity involves the analysis of a target construct in relation to its alternatives or cognates" (Shadish et al., 2002 p. 364) to show that the target construct is distinct from others. Campbell and Fisk (1959, p. 84) explain further:

[1] Straub, Boudreau, and Gefen's (2004) work covers in detail many forms of validity including nomological, predictive, concurrent, factorial, manipulation, and statistical conclusion validity. They also cover important types of reliability including split-half, test-retest, inter-rater, unidimensional, and alternate forms of reliability.

> *When a construct is proposed, the proponent invariably has in mind distinctions between the new dimension and other constructs already in use. One cannot define without implying distinctions, and the verification of these distinctions is an important part of the validation process.*

Discriminant validity offers statistical support that a theoretical distinction exists between the constructs of interest. Once reliability and validity can be shown, researchers can begin to have confidence in using the measure. Nevertheless, good researchers will always be circumspect and make efforts to check reliability and validity and, if possible, improve reliability and validity in each study they undertake.

2.1 Establishing Discriminant Validity with Highly Correlated Constructs

Normally, highly correlated constructs present a statistical challenge in analyzing models and can threaten the validity of the analysis results. For example, multicollinearity occurs when two predictors in a regression equation are highly correlated. When this happens, regression coefficients are unreliable and results can be difficult to interpret. Interestingly, in research situations where strong predictors are sought after to explain variance, researchers are often criticized for including multiple highly correlated constructs in their models. Thus, some researchers work to avoid highly correlated predictors. Here, we use a sample dataset that has a single instance of such a case. However, it is not uncommon to have multiple instances in doing such research. We show that using highly correlated constructs together in a model is appropriate conceptually speaking and can be statistically acceptable with appropriate techniques and clear theoretical support.

A possible difficulty with assessing discriminant validity arises when a statistical analysis does not support discriminant validity and yet the theoretical framing does support the distinction between constructs. In some cases, this problem represents a measuring problem that should be addressed as part of the content validity assessment step. In some other cases, it represents lack of support for distinction. The researcher must then assess the appropriateness of the analysis and the results and weigh the information against the theory. In such a circumstance, researchers have three alternatives to consider: 1) the constructs are not distinct, 2) there is a legitimate direct association between the two constructs, and 3) there is a mutual association with another construct. The first alternative (that the constructs are not conceptually distinct) must be assumed if conceptual and empirical results cannot support an alternative explanation. Making such assessments requires a level of understanding regarding the underlying theory, observed data, and statistical analysis techniques.

The second alternative is a direct association between constructs. This situation must appeal to the theoretical understanding of the two constructs and be supported with empirical results. For example, research had shown that the constructs of motivation and intelligence are highly correlated (Muir & de Charmsm 1975). As a result, two schools of thought emerged regarding this topic. The first school of thought advocated the existence of two highly correlated, yet distinct, latent constructs (Burt & Williams, 1962). Conversely, a second school of thought emerged that did not support this view and claimed that motivation and intelligence are two different labels for essentially the same latent construct (Loretan, 1965). A more recent example of this type of conversation can be found with regard to the constructs of emotional intelligence and motivation. Some researchers support that motivation is integral to emotional intelligence (Goleman, 1995; Goleman, 1998), and others support their distinctness (Mayer & Salovey, 1997). This conversation continued being discussed in the literature nearly a decade later (Christie, Jordan, Troth, & Lawrence, 2007). Since neither side can offer irrefutable evidence to support their view, it is left to each researcher to decide which school of thought they accept.

The third alternative is a mutual association with another construct. There may be a third variable that, if included, could explain an apparent relationship between two constructs and aid to their distinction. Identifying such a construct may be done by an appeal to theory or an appeal to data in identifying a nuisance variable. A nuisance variable is not of theoretical interest in a relationship yet affects the relationship. Common method bias may be an example of an outside factor mutually influencing an association between constructs. Most of these cases can be statistically controlled by parsing out the factor responsible in the mutual association.

Considering these three alternatives underscores the importance of theoretical understanding and support for discriminant validity analysis. With this focus in mind, we illustrate the analysis conducted during an instrument development process. During this illustration we highlight some alternative discriminant validity analysis techniques along with their respective strengths and weaknesses.

3 Illustration with Sample Data

We collected the data we use in this example with the intent of developing an instrument of newly applied constructs extended from psychology literature and instantiated into an IS context. We used two separate and independent samples in this process. We performed an exploratory data analysis on the first sample, while we used the second sample to cross-validate the results of the first sample (Byrne 2006). We developed the survey instrument by undergoing the following process: item generation, exploratory factor analysis, cross-validation of the exploratory analysis with a second sample, and the assessment of the instrument's reliability and validity.

3.1 Construct Definition

To allow some theoretical understanding, Figure 1 presents a research model and Table 1 presents the definitions of these constructs. The ultimate dependant variable of this model is a specific type of behavioral intention, the intention to form a long-term business-to-consumer relationship with an organization. This research model is a variant of the theory of reasoned action, with various attributes leading to a belief about taking part in a relationship, which ultimately affects an individual's intention to take part in said relationship. We extended the immediate predictor and antecedents of the research model from an interpersonal relationship theory called stage theory (Levinger, 1980).

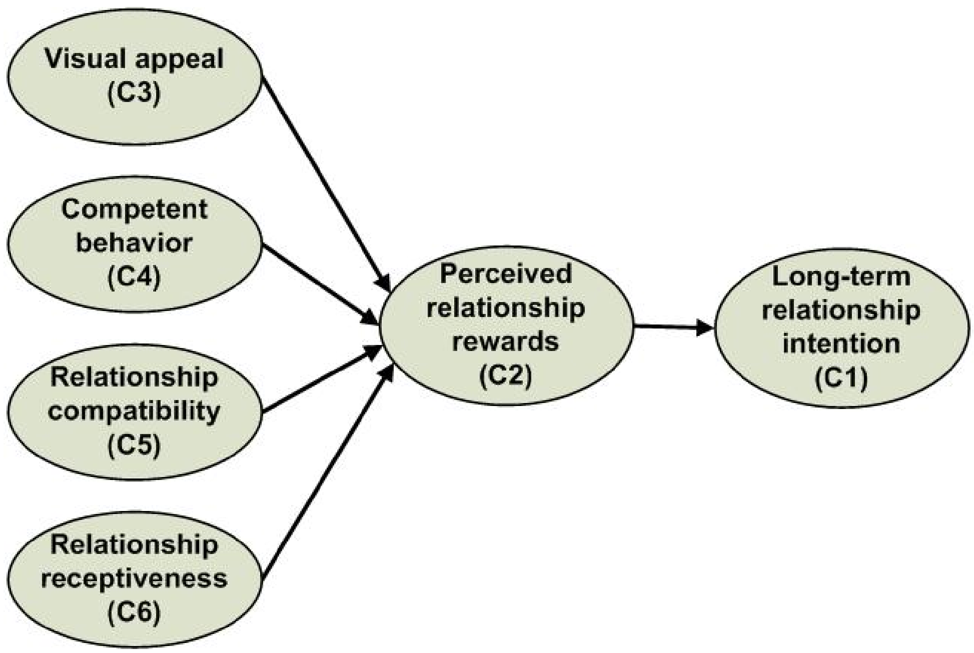


Figure 1. Example Research Model

Table 1. Construct Definitions Related to the Research Model (see Appendix A for Items)

Construct	Definition
Long-term relationship Intention	The specific behavioral intention to engage in a long-term customer (business-to-consumer) relationship with an online firm
Perceived relationship rewards	Perceptions of overall possible benefits from interactions with a Web-based organization in an online B2C relationship
Visual appeal	Overall perceptions of a website's aesthetics and appearance
Competent behavior	Perception of competence of the IS interface and its functionality
Relationship compatibility	Perception that the website content communicates values and beliefs that are compatible with the values and beliefs of the user
Relationship receptiveness	Perception of a company's desire to enter into a customer relationship

Research on attraction between humans has found that the primary factor for engaging in a relationship is the overall perceived rewards of the relationship (Aronson & Linder, 1965). Stage theory offers the following determinants of perceived rewards in the beginning stages of a relationship: good appearance, competent behavior, compatibility, and a level of responsiveness (Huston & Levinger, 1978, Levinger 1980). Recent research has extended and operationalized these constructs (see Table 1) in the context of B2C relationships (Campbell, Wells, & Valacich, 2009).

3.2 Content Validity

To ensure content validity, three researchers generated and subjectively analyzed potential items for each of the constructs in the research model. The researchers appraised the pool based on their knowledge of the theoretical domain. This process resulted in approximately 30 potential items for each construct. We expected to drop many of these items during the psychometric assessment. We grouped these items and included them in a survey along with other previously validated measures.

The first survey included 395 undergraduate college students. Because these measures are intended to assess the initial perceptions of perspective customers toward e-commerce firms based on websites, we gave the subjects a scenario search task on existing e-commerce websites prior to filling out the survey. We eliminated participants that reported to have visited the website previously from the dataset. The average age of the subjects was 21.91 years, and there were 218 (55.2%) males. Students received course credit—approximately 1 percent of their final grade—for participating in the survey.

3.3 Survey 1: Exploratory Analysis Pre-Test Item Assessment

Subsequently, using SPSS 11.0 based on maximum likelihood estimation, we conducted an exploratory factor analysis (EFA) to identify the most promising items. We identified eight items for each construct that all loaded at higher than .7 on their assigned factor (See Appendix B for EFA results). We then included the items in a structural equation modeling (SEM) measurement model using EQS 6.1. We used modification indices to refine the number of items per construct. The procedures for selecting these items on modification indices used the following guidelines: 1) Deletion of items with the highest modification indices or with error terms associated with the highest modification indices, 2) Items with highest modification indices were not deleted in the case that the construct would then be represented by fewer than three items, and 3) Items with the highest modification indices were not deleted once the measurement model attained an acceptable fit. We then used modification indices to identify items that were highly correlated to factors in which they were not assigned and to eliminate them from consideration. Next, we collected a second sample to cross-validate the findings of this exploratory analysis and assess the reliability and validity.

3.4 Survey 2: Cross-validation

The data gathering procedure for this second sample was similar to the first. We revisited the survey based on the results from the first sample. We limited the survey to the items retained after the EFA in study 1. The second sample comprised 275 responses. The average age of the subjects was 22.21, and there were 180 (65.5%) males.

3.5 Measurement Model Assessment

Choosing the appropriate fit statistics is an important step in covariance-based SEM. Covariance-based SEM is not robust to high levels of multivariate kurtosis or non-normality (Bentler, 2005; Byrne, 2006; Curran, West, & Finch, 1996; West, Finch, & Curran, 1995). Unfortunately, data collected via survey instruments can have high levels of multivariate kurtosis, which proved to be the case in our sample. Because we assigned participants to different treatments, one could expect that responses would not follow a normal distribution. The multivariate kurtosis for this measurement model was 54.36, which exceeds the recommended parameters of being no greater than 5 (Bentler, 2005; Byrne, 2006). Evaluation using the chi-square (χ^2) statistic (or variants of the χ2 statistic) may not be adequate under these conditions (Hu, Bentler, & Kano, 1992). Therefore, corrected fit statistics have been found to be more appropriate (Hu et al., 1992). Satorra and Bentler (1988) developed a scaling correction for the χ^2

statistic that has been shown to be most reliable (Curran et al., 1996; Hu et al., 1992; Satorra & Bentler, 1988). In this paper, we evaluate a model's fit based on the Satorra-Bentler scaled (S-B χ^2)[2] fit indices.

We selected the following fit statistics for the purposes of this study: the robust comparative fit index (CFI), the robust root-mean-square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). The criteria we used to evaluate model fit were: CFI values must be .95 or higher, SRMR values must be .08 or lower, and the RMSEA values must be .08 or lower (Hu & Bentler, 1999)[3]. The measurement model below (see Table 2) complied with these thresholds and, thus, demonstrated good fit.

Table 2. Measurement Model: Standardized Loadings (All Loadings p<.0001), Composite Reliabilities and Fit Statistics

Items	Standardized loadings	Composite reliabilities	Alphas	Items	Standardized loadings	Composite reliabilities	Alphas
Competent behavior				**Long-term relationship intention**			
1	0.882			1	0.862		
2	0.894			2	0.902	0.899	0.894
3	0.920	0.927	0.935	3	0.828		
4	0.716			**Relationship compatibility**			
5	0.901			1	0.895		
Visual appeal				2	0.819		
1	0.967			3	0.843	0.927	0.942
2	0.973	0.979	0.980	4	0.926		
3	0.970			5	0.915		
Perceived relationship rewards				**Relationship receptiveness**			
1	0.876			1	0.896		
2	0.822	0.898	0.898	2	0.864	0.938	0.938
3	0.893			3	0.884		
				4	0.912		
Fit statistics							
χ^2/df	682.211/215			SRMR	.052		
S-B χ^2	436.2805			RMSEA	.061 (.053,.069)		
CFI	.959						

3.6 Reliability Analysis

We conducted a reliability analysis for these constructs using the Cronbach alpha and the composite reliabilities (Werts, Linn, & Joreskog, 1974). Cronbach's alpha is "a measure of reliability or consistency of the items in an index." (Vogt, 1999, p. 64) The literature recommends that the Cronbach alpha should be above 0.70 (Nunnally & Bernstein, 1994) and composite reliabilities (Hair, Anderson, Tatham, & Black, 1998). As Table 2 shows, all these recommended thresholds were met and support the reliability of these measures.

3.7 Convergent Validity

One can use the factor loadings as a measure of the degree of generalizability found between each observed variable and its factor. These loadings should exceed the threshold of 0.707 to demonstrate convergent validity (Chin, 1998; Hair et al., 1998; Segars, 1997). Also, one can use the average variance

[2] The definition for the Satora-Bentler scaled statistic can be found in the appendix of Hu et al. (1992, p. 351): "It is difficult to summarize verbally and succinctly this technical literature, but it has been shown that asymptotic optimality and correct standard errors of factor loadings can be obtained under normal-theory methods when the common factors are not normally distributed and the unique factors have a multivariate normal distribution and hence the observed variables are also nonnormal.".
[3] Research indicates that LTI fit statistic is also useful because it is not affected by sample size (Vandenberg & Lance, 2000).

extracted (AVE) above .50 (Fornell & Larcker, 1981). The AVE of our constructs was as follows: long-term relationship intention (0.747), perceived relationship rewards (0.747), visual appeal (0.941), competent behavior (0.808), relationship compatibility (0.760), and relationship receptiveness (0.791)[4]. The factor loadings (see Table 2) and the AVE of each construct indicate compliance with these standards. As such, we conclude that convergent validity has been demonstrated.

4 Discriminant Validity

Assessing discriminant validity in this sample data provided mixed results. In Section 3, we discuss how theory is used in the construct definition and how it plays an important role in analyzing empirical results. Additionally, theory is used in understanding the distinction between the appropriate use of certain analysis techniques such as EFA and confirmatory factor analysis (CFA). EFA is data driven and conducted to discover patterns in the data that suggest latent variables in a dataset. In contrast, CFA tests theories and hypotheses about the factors one expects to find (Vogt, 1999). Thus, CFA is more theory driven than EFA. Note that EFA can be used for theory-driven analysis and CFA for data-driven analysis; however, it is important to understand the difference between these analysis techniques. Therefore, we emphasize the importance of not only understanding analysis output but also the purpose of the analysis technique and the limitations of what the output describes about a given relationship between constructs. Such an understanding will allow a richer understanding of the relationship between constructs.

We first highlight, in the interest of completeness, some discriminant validity techniques that we did not use in our data collection or analysis (i.e., Q-sorting, multi-trait multi-method approach, and hierarchical cluster analysis). We then use our sample data to focus on the strengths and weaknesses of some commonly used discriminant validity analysis techniques.

4.1 Q-Sorting

Q-sorting is a long-established method for analyzing subjective content in a theoretically grounded and quantitative method (Brown, 1986). Thomas and Watson (2002) thoroughly explain the application of Q-sorting for MIS research and summarize this method as follows:

> *Q-sorting consists of “modified rank-ordering procedures in which stimuli are placed in an order that is significant from the standpoint of a person operating under specified conditions.” (Brown 1980 p. 195) It results in the captured patterns of respondents to the stimulus presented, a topic on which options vary. Those patterns can then be analysed to discover groupings of response patterns, supporting effective inductive reasoning (Stephenson 1979).* (Thomas & Watson, 2002, p. 141)

Straub et al. (2004 p. 390) point to various researchers that use Q-sorting as an “innovative means of verifying discriminant validity” (Moore & Benbasat, 1991; Segars & Grover, 1998; Storey, Straub, Stewart, & Welke, 2000). However, Thomas and Watson (2002, p. 152-153) outline some of these same specific cases (e.g., Moore & Benbasat, 1991; Segars & Grover, 1998) and warn of a potential weakness in using this type of analysis for the purposes of establishing validity because it violates a key assumption:

> *Violation of the forced-distribution requirement during data collection invalidates the principles of psychological significance and choice equilibrium underlying self-reference and leads to questions about the applicability of Q-method’s theoretical foundation during data analysis (Brown 1980).* (Thomas & Watson, 2002, p. 153)

Another inherent weakness of the Q-sort is the influence of outside factors that are commonly known to confound human subjectivity (e.g., social desirability). However, especially considering our focus here on theory’s role in these analyses, one must consider it a strength that subjectivity in these assessments are valued and quantitatively analyzed using factor analysis. An additional strength of this method is in using induction or deduction to produce insights in the final steps of Q-sorting. Therefore, we submit that using this method in conjunction with other confirmatory methods represent a complementing approach to discriminant validity analysis supported by past literature (Lewis et al., 2005; MacKenzie et al., 2011). Table 3 summarizes these strengths and weaknesses.

[4] In many papers, AVE’s, squared AVE’s (often on a diagonal), and correlations are reported in the same table (see Table 9 below).

Table 3. Strengths and Weaknesses of Q-sorting for Discriminant Validity

Strengths	• Initial assessment of items based on conceptual definitions • Introduces quantitative assessments of subjective assessments • Allows for post-hoc induction based on results and conceptual understanding • Not specific to any one method of hypothesis testing (e.g., ANOVA, regression, SEM, etc.)
Weakness	• Results may be affected by errors in subjectivity (e.g., social desirability) • Research suggests that using this method for discriminant validity may be a violation of the forced-distribution requirement that is assumed in Q-methods

4.2 Multi-trait Multi-method

The multi-trait multi-method (MTMM) approach is well established in social science (Campbell & Fiske, 1959) and IS research (Straub, 1989). The MTMM is a:

> *highly formal approach to validation which involves numerous comparisons of correlations and correlational patterns. Percentages smaller than chance of violations of convergent and discriminant validity conditions in the matrix of trait (or item) and method correlations indicate that the methods are equally valid.* (Straub et al., 2004, p. 391)

The MTMM, compares correlations across methodological approaches and similar constructs. If correlations converge as theorized (i.e., higher correlations for like constructs across methods), discriminant validity is supported. However, Straub et al. (2004, pp. 391-392) also summarize the weaknesses of this approach as follows:

> *Problems with MTMM are legion. Bagozzi (1980) and Bagozzi and Phillips (1982) argue that counting violations in a correlation matrix is an arbitrary procedure that can lead to incorrect conclusions. If a researcher gathered data via more than one method, Bagozzi (1980) shows how SEM can be used to examine method versus trait variance as well as other validity properties of the entire MTMM matrix of correlations. "MTMM's requirement for gathering of data through at least two "maximally different methods" (Campbell and Fiske 1959 p. 83) places such a heavy burden on researchers that they may be shying away from it. In fact, no matter how much the community wishes to have valid instruments, it is possibly overmuch to ask researchers to engage in this level of validation at an early stage in a research stream.*

Although the MTMM's weaknesses are well known, it remains a widely accepted method for establishing the validity of measurement instruments. The method's strength is that it does not over rely on a single method of measurement and, thus, guards against common method bias, which has been shown to be a significant issue in IS research (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Table 4 summarizes these strengths and weaknesses.

Table 4. Strengths and Weaknesses of MTMM for Discriminant Validity

Strengths	• Robust to common method bias due to use of "maximally different" methods • Highly used and respected since Campbell and Fiske (1959)
Weakness	• Research has shown that interpretation of correlations may lead to incorrect conclusions • Multiple methods may not be available, especially for newly identified variables

4.3 Hierarchical Cluster Analysis

Hierarchical cluster analysis is a process of classifying groups out of varying objects. Since objects can represent any number of things such as subjects, patients, stimuli, concepts, and variables (Kloot, Spaans, & Jeiser, 2005), this type of cluster analysis has been used in many different types of disciplines ranging from medical science and biology to marketing research. There have been many different proposed methods of clustering in the area of social sciences. These methods have largely been grouped into two types: agglomerative (which begins with the objects and successively groups them together) and divisive (which begins with the root and proceeds to split them into clusters) (Aldenderfer & Blashfield, 1984; Everitt, 1993; Gordon, 1999). The objects and clusters are often depicted in a dendrogram that also visually communicates a metric of distance between the objects and the linkage between items. To support discriminant validity, the distance between like items should be small compared to those of other constructs. However, the assumptions and boundaries associated with various metrics of distance and linkages are still a highly studied and disputed area of research (e.g., Kloot et al., 2005).

Two weaknesses of the hierarchical cluster analysis are worth mentioning. First, one criticism of cluster analysis is that there seems to be no one discipline responsible for advances in cluster analysis (Punj & Stewart, 1983). Therefore, most applications are extended from various contexts with several different procedures, analysis techniques, and assumptions, which makes it difficult to come to agreement on the most appropriate type for any one area of application or make comparisons (Kloot et al., 2005). Second, Punj and Stewart (1983, p. 145) note another weakness by stating that "a cluster solution will be reached even when there are no natural groupings in the data". Therefore, a need still exists to further assess the reliability of these types of results with another analysis technique. Given these challenges, these techniques do not appear regularly in IS research. Table 5 summarizes these strengths and weaknesses.

Table 5. Strengths and Weaknesses of Hierarchical Cluster Analysis for Discriminant Validity

Strengths	• Provides opportunity for subjectively assessing items based on a conceptual definition (i.e., not purely data driven) • May result in identifying sub-constructs • Not specific to any one method of hypothesis testing (e.g., ANOVA, regression, SEM, etc.)
Weakness	• No single accepted method has been identified, which has resulted in competing perspectives on analysis and interpretation • May result in unreliable results.

4.4 Strengths and Weaknesses of Commonly Used Post-Hoc Discriminant Validity Analyses

Lewis et al. (2005) propose a construct development methodology that distinguishes between three different stages of development: 1) domain, which comprises content analysis, 2) instrument, which comprises item screening, and finally 3) measurement properties, which comprises exploratory and confirmatory analysis. Similar methodologies have recently been proposed (MacKenzie et al., 2011). Two of the above (i.e., Q-sorting and hierarchical cluster analysis) are established discriminant validity techniques that focus on the first two stages of development and are not the focus of this paper. We use the following exercise to demonstrate the strengths and weaknesses of some commonly used analysis techniques in IS research that assess measurement properties (e.g., confirmatory) in order to highlight the role of theoretical understanding in interpreting the results. Of specific focus are the following discriminant validity analysis techniques: correlation-based, average variance extracted (AVE) comparisons and $\chi 2$ comparisons.

4.5 Correlation-based Analysis

The correlation-based analysis is a discrimination test based on examining the correlation between two factors (e.g., Carlson, Kacmar, & Williams, 2000; Kline, 2005; McKnight, Choudhury, & Kacmar, 2002). Interpreting these correlations is a subjective process that is at the researcher's discretion. The primary goal of this analysis technique is to identify any correlations that support two constructs as being (in practice) the same. Kline (2005) offers a rule of thumb that the upper limit for correlations between factors should be .85. Kline argues that items measuring two constructs that correlate at this level can hardly be expected to represent two distinct constructs. The .85 threshold, however, is not positioned as an explicit rule because it is not supported with any empirical evidence; this rule-of-thumb is simply presented as a suggested reference point. These rules-of-thumb and de facto standards do provide some practical guidance but are also considered to be inherently limited:

> *Validity rules of thumb are pragmatic measures indicating patterns of behavior that are acceptable within a scientific community. There is no recognized means of verifying the truth of*

> *such heuristics, other than through tradition or evaluation of best of breed practice. It is traditional, for example, to accept a p-value of .05 in SEM (Joreskog and Sorbom 1983), just as the .01 and .05 thresholds are accepted heuristics in linear regression (Neter, et al. 1990). As with first generation regression models, there is no mathematical or other means for establishing these levels (Nunnally 1967, Nunnally 1978, Nunnally and Bernstein 1994). Nonetheless, rules of thumb are desirable because of their practicality, enabling researchers to utilize them as de facto standards.* (Gefen et al., 2000, pp. 42-43)

Consistent with this line of thinking, we perceive these rules-of-thumb and de facto standards as being helpful yet, at the same time, inherently limited.

The limitation of this type of discriminant validity check is that there is no statistical test to determine if constructs are discriminant. This type of analysis requires each researcher to set their own threshold and interpret the data themselves. This presents an obvious problem because researchers often hold different standards for discriminant validity. For instance, some scholars contend that correlations above .6 (more than 36% shared variance) could be too high (e.g., Carlson et al., 2000; Loehlin, 2004, p. 99; McKnight et al., 2002). Thus, the weakness with correlation-based analysis is that it may be difficult for researchers to agree on an acceptable level of correlation. This is especially true when considering the role of theory. Some theoretical perspectives may support a high correlation and others may not. Therefore, the interpretation becomes subjective and context specific. Table 6 summarizes the strengths and weaknesses of this method.

Table 6. Strengths and Weaknesses of Correlation-Based Assessment of Discriminant Validity

Strengths	• Allows for the theoretical perspective to guide interpretation • Compares all possible relationships within a pool of constructs • Not specific to any one method of hypothesis testing (e.g., ANOVA, regression, SEM, etc.)
Weakness	• No common threshold available that all researchers agree on • Allows for misinterpretation • Allows for multiple interpretations

Table 7 reports the correlations comparison that we used to initially assess the discriminant validity of the constructs in the research model. This analysis indicates that the discriminant validity between perceived relationship rewards and long-term relationship Intention may be in question. However, according to the theoretical perspective shown in the research model, the primary factor for engaging in a relationship is the overall perceived rewards of the relationship (Aronson & Linder, 1965). Therefore, these factors are expected to have a high correlation. This high correlation may be theoretically justifiable, and it may be necessary to further investigate the discriminant validity between these two constructs.

Table 7. Factor Correlations for Survey 2

	C1	C2	C3	C4	C5	C6
Long-term relationship intention (C1)	1.000					
Perceived relationship rewards (C2)	0.915	1.000				
Visual appeal (C3)	0.689	0.75	1.000			
Competent behavior (C4)	0.637	0.771	0.73	1.000		
Relationship compatibility (C5)	0.696	0.764	0.626	0.576	1.000	
Relationship receptiveness (C6)	0.632	0.742	0.661	0.726	0.658	1.000

4.6 AVE Comparison Analysis

Another discriminant validity analysis technique is AVE comparison. The AVE comparison technique is growing in popularity for IS research. This is especially true because it was developed for use in conjunction with PLS-based SEM, which has been noted to provide various advantages over covariance-based SEM (e.g., more appropriate for exploratory analysis and more robust to smaller sample size). Conceptually, the AVE denotes the variance explained in a factor by its items. AVE comparison analysis for discriminant validity is used when the AVE is calculated for each construct (see Appendix C) and compared to the squared correlation (i.e., variance explained by another construct) between two constructs (Gefen & Straub, 2005; Gefen et al., 2000). If the AVE is larger than any squared correlation, then it is said to show discriminant validity because the items account for more variance than any variance explained by a correlation with another construct (Anderson & Gerbing, 1988; Segars, 1997).

The difficulty with this method lies in the comparison made between two different sources of variance accounted for (VAF). The first source of VAF is the constructs own assigned items. The second source of VAF is the correlation with another latent variable. This is not a direct comparison (e.g., apples-to-apples) because one measure of VAF is at the indicator level (although aggregated) and the other is at the construct level. To demonstrate the possible complications with this comparison, consider the following examples.

First, consider a common method bias inflating the correlation between the constructs. Depending on the different ways in which the questionnaire is constructed, there could be an inflation of variance among the items of each construct, among the constructs, or among both. An AVE analysis does not correct for such cases. A second example could include limitations in the measurement of constructs such as the Mach IV. The Mach IV scale has been largely criticized for low reliability with coefficients ranging from .59-.88 (Moss, 2005) in the measurement of the Machiavellian personality trait. However, the Mach IV scale is largely accepted in its use due to the complexity of the construct definition. Consider the result of a discriminant validity analysis that indicates a higher squared correlation with the Mach IV. Does that support the conclusion that the constructs were indiscriminant considering the limitations of the Mach IV's own indicators? Of course, in such a case, the limitations of the Mach IV would be of most concern. However, even in situations with correlations to valid and reliable measures, items never represent a latent construct perfectly exhibiting unique variance (i.e., associated disturbances). In both these cases, we highlight important variables that effect the measurement of correlations differently than the item loadings used in AVE calculation.

This method constitutes a sufficient, but not necessary, condition for discriminant validity. If the AVE is higher than a squared correlation, then it supports discriminant validity. If the AVE is not larger, however, one cannot assume that discriminant validity is unsupported. Table 8 summarizes the strengths and weaknesses of the AVE comparison method for establishing discriminant validity.

Table 8. Strengths and Weaknesses of AVE Comparison for Discriminant Validity

Strengths	• Established de facto standard • Accounts for explanatory power of a construct's items • Correlational data are included in the analysis, but interpretation is more objective than a raw interpretation of a correlation • Can be used with PLS-based SEM and covariance-based SEM
Weakness	• No significance test • Comparison represents a sufficient but not necessary condition for discriminant validity • Comparing the VAF of constructs items and the VAF of a correlation to another latent variable is not an apples-to-apples comparison

Table 9 shows the AVE comparison for survey 2's data. These results indicate that the discriminant validity between perceived relationship rewards and long-term relationship intention may be in question. This becomes more interesting when considering the results of the following analysis technique, especially considering the bias of the AVE comparison with correlational data.

Table 9. EQS Estimated Squared Correlations and (AVE)*

	C1	C2	C3	C4	C5	C6
Long-term relationship intention (C1)	**0.747**					
Perceived relationship rewards (C2)	0.837	**0.747**				
Visual appeal (C3)	0.475	0.563	**0.941**			
Competent behavior (C4)	0.406	0.594	0.533	**0.808**		
Relationship compatibility (C5)	0.484	0.584	0.392	0.332	**0.760**	
Relationship receptiveness (C6)	0.399	0.551	0.437	0.527	0.433	**0.791**

* AVE values are shown in bold along the diagonal, and the others are the squares of those in Table 7.
The AVE statistic is given in PLS outputs, and can be manually calculated using covariance-based SEM as $(\Sigma\lambda_i^2)/[(\Sigma\lambda_i^2) + \Sigma_i Var(\varepsilon_i)]$ where λ_i is the indicator loading and $Var(\varepsilon_i) = 1 - \lambda i^2$).

Often, a researcher, in the face of new evidence against discriminant validity such as this AVE analysis, would return to the items and subjectively assess them to find out “what went wrong”. Such an item analysis is not consistent with a CFA approach because discriminant validity is a question for the construct level. A subjective post-hoc analysis of these items could reveal that one of the items for long-term relationship intention may conceivably measure some aspects of perceived rewards. Based on this new point of view, a) the instrument could be rejected by citing Anderson and Gerbing (1988, p. 415), who state: “measurement models that contain correlated measurement errors or that have indicators that load on more than one estimated construct do not represent unidimensional [which is not normally assessed in discriminant validity analysis] construct measurement (Gerbing & Anderson, 1984)”, or b) another perspective could be set forth that is often practiced in theory-driven research (e.g., Burns, Boe, Walsh, Sommers-Flanagan, & Teegarden, 2001; Vallerand & Richer, 1988) to accept such cross-loadings if the loadings are significant and the wording of the items are consistent with the construct definition. Neither stance is correct in all cases. Therefore, a clear understanding of the theoretical perspective and the strengths and weaknesses of the empirical analysis techniques is needed.

4.7 χ^2 Comparison Analysis

χ^2 comparison analysis compares the χ^2 values between fixed and free solutions for each pair of constructs being assessed in a measurement model using SEM analysis (see Appendix D) (Anderson & Gerbing, 1988b; Segars, 1997; Straub et al., 2004). The positive aspects of this type of discriminant analysis are that it is not reliant on correlational data, de facto standards, or rules of thumb. Of the discriminant validity analysis techniques discussed in this paper, the χ^2 comparison analysis is the only one that offers a statistical significance test. The test’s weakness is that it is highly influenced by sample size (Byrne & Stewart, 2006; Chen, Sousa, & West, 2005; Cheung & Rensvold, 2002) and non-normality (West et al., 1995).

As the sample size increases, the likelihood of a significant difference increases since the χ^2 statistic is directly related to sample size (χ^2 = (N-1)FGLS):

> *Recently, researchers (Cheung & Rensvold, 2002; Little, 1997; Marsh, Hey, & Roche, 1997) have argued that the $\Delta\ \chi^2$ value is as sensitive to sample size and non-normality as the chi-square statistic itself, thereby rendering it an impractical and unrealistic criterion.* (Byrne & Stewart, 2006, p. 305)

Given the weakness of the χ^2 comparison analysis in using large sample sizes, one should understand the research regarding power analysis in SEM. Note that, even with a justifiable sample size in terms of power, the sample may still be too large for the χ^2 comparison analysis. The rules-of-thumb for sample size help give perspective to researchers’ confidence in these types of analysis.

One such rule-of-thumb is to have five to ten subjects per estimated parameter in the model, which must be considered a minimum (if not too few) from a power perspective. If there are high levels of non-normality or missing data, it is justifiable to have 20 subjects per estimated parameter. However, in terms of power, there should be no upper limit considered. An estimated parameter is a parameter in the SEM model that is neither fixed nor observed (e.g., covariances, variances, disturbances, errors, path weights).

Therefore, the more items (observed variables), relationships, and constructs that are represented in a model, the more parameters are estimated. In such cases, more subjects are needed to attain adequate power.

We believe that the rules-of-thumb we summarize here are justifiable as comparison points for determining if the sample size is defensible for a χ^2 comparison analysis. Five to ten subjects per estimated parameter are needed, as a minimum, if the data is normally distributed and if there is no missing data. Twenty subjects per estimated parameter are needed, as a minimum, if the data has levels of non-normality or missing data points. However, further simulation research is needed for empirical standards to be offered. Table 10 summarizes the strengths and weaknesses of this method.

Table 10. Strengths and Weaknesses of χ^2 Comparison Discriminant Analysis

Strengths	• Significance test is available • Not reliant on correlational data
Weakness	• The statistic is skewed by large samples size. Therefore, small and practically irrelevant relationships could be determined statistically significant • The statistic is skewed by non-normality • Limited to use with covariance-based SEM that provides χ^2 statistic

We used the χ^2 discriminant comparison analysis to examine the discriminant validity of the research model (see Table 11). The results indicate that every combination of factors results in a significant decrement in fit compared to the original model. χ^2 difference test for 1 df at the p<.001 requires a difference of 10.828 in the χ^2 values (Baker, 2000). As such, one can conclude that these constructs show discriminant validity using this method.

Table 11. χ2 Discriminant Analysis

Model	df	χ^2 Value		Model	df	χ^2 Values
Original	**215**	**682.211**		Relationship receptiveness, perceived relationship rewards	216	732.494
Relationship compatibility, competent behavior	216	700.806		Relationship receptiveness, visual appeal	216	745.879
Relationship receptiveness, relationship compatibility	216	707.408		Relationship receptiveness, competent behavior	216	748.925
Relationship receptiveness, long-term relationship intention	216	709.221		Long-term relationship intention, visual appeal	216	750.203
Relationship compatibility, long-term relationship intention	216	710.881		Perceived relationship rewards, competent behavior	216	754.730
Long-term relationship intention, competent behavior	216	717.296		Perceived relationship rewards, visual appeal	216	774.020
Relationship compatibility, perceived relationship rewards	216	723.061		Visual appeal, competent behavior	216	786.134
Relationship compatibility, visual appeal	216	724.014		Long-term relationship intention, perceived relationship rewards	216	794.689

4.8 Assessing the Distinction between the Constructs

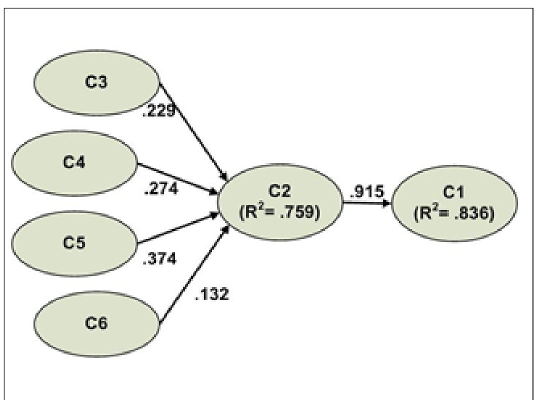
Some of the preceding data analysis techniques do not support that perceived relationship rewards and long-term relationship intention are discriminant with the exception of the χ^2 discriminant analysis, which is the only test that offers a significance test and is not driven by correlational data. Assuming a lack of definitive support either way, it may be necessary to conduct further analysis based on the presumed theoretical perspective or return to the starting point and develop further refined items to measure these constructs. Given the theoretical perspective presented in the research model (see Figure 1), there are four possible ways to work with this situation dealing with discriminant validity (see Table 12). We explore each of these options to further investigate the proposed theoretical relationship between constructs long-term relationship intention and perceived relationship rewards.

Table 12. Options for Dealing with Discriminant Validity Between Long-Term Relationship Intention and Perceived Relationship Rewards

Option A	Combine the factors and test the model with all the manifest variables from both perceived relationship rewards and long-term relationship intention to one single construct.
Option B	Delete long-term relationship intention from the model and have perceived relationship rewards be the ultimate exogenous variable.
Option C	Delete perceived relationship rewards from the model and have long-term relationship intention as the ultimate exogenous variable.
Option D	Leave the model as proposed and tentatively accept the high correlation between the factors as the estimated relationship between these two constructs.

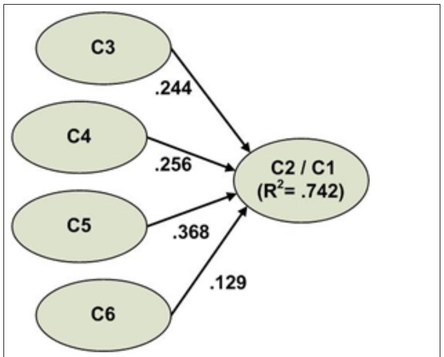
Each of these options is relatively defensible given the results of this data. Option A concedes that the two constructs are not conceptually different and, therefore, can be modelled as a single construct represented by manifest variables originally designed to represent distinct constructs. Option B also concedes that the constructs are not distinct and eliminates the intention construct and associated manifest variables from the model. Option C is similar to option B but differs in that it illuminates the rewards construct. Finally, option D conceptually supports the existence of two distinct constructs and accepts the high correlation as representative of the theorized relationship. However, the question of which option is most appropriate remains. We compare each of these options and use various statistical analyses to identify the most appropriate option given the theoretical framing in this particular study.

Option A suggests that long-term relationship intention and perceived relationship rewards are too similar. The assumption here is that the constructs are so similar that, from a pragmatic perspective, they can be treated as the same construct. Therefore, all of the items could be combined. To test this assumption empirically, we compared option A and option D with an appeal to a fit statistic that is designed for such a comparison. We used the Akaike's information criteria (AIC) to compare two or more models with the smaller values representing a better fit (Byrne, 2006; Hu et al., 1992). Raftery (1995) indicates that drops in the AIC greater than ten can be considered quite significant. To make this comparison, structural models were fit for these two options (see Figures 2 and 3).



AIC	252.757	SRMR	.052
χ^2	681.757	CFI	.959
S-B χ^2	435.660	RMSEA	.061 (.053,.069)
All paths significant at $p<.05$			

Figure 2. Structural Model: Standard Regression Weights, Variance Explained, and Fit Statistics for Option D



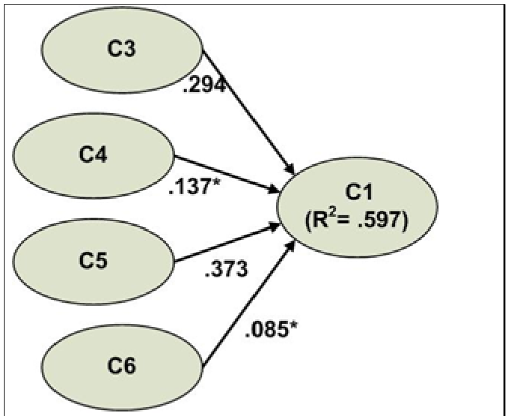
AIC	288.294	SRMR	.049
X2	728.294	CFI	.954
S-B χ^2	465.751	RMSEA	.064 (.056,.072)
All paths significant at p<.05			

Figure 3. Structural Model: Standard Regression Weights, Variance Explained, and Fit Statistics for Option A

The results of these structural models show that both options meet the fit thresholds. A review of the AIC for Figures 2 and 3 show that the model with separate factors (Figure 2) illustrates better fit than modelling the factors combined (Figure 3). These results support option D over option A.

Option B assumes that there is no distinction of long-term relationship intention and perceived relationship rewards, and that long-term relationship intention can, therefore, be deleted from consideration. Option B is not theoretically supported. The proposed model is intended to predict a behavioral intention that past research has shown to be a strong predictor of actual behavior. Such a link has not been established between the rewards construct and actual behavior. Therefore, eliminating the intention construct leaves this research model with little theoretical contribution to offer. This option, therefore, shifts the focus from theoretically driven to data-driven research. However, option B should be recognized as a viable option to address the statistical issues raised in these analyses.

Options B and C are based on a similar theoretical assumption; specifically, that there is no distinction between long-term relationship intention and perceived relationship rewards. And each option proposes one or the other construct, together with the associated manifest variables, be deleted from the research model. Figure 4 summarizes the details of a structural model fit for option C. The results summarized in Figure 4 do not support that perceived relationship rewards and long-term relationship intention are highly correlated and, thus, does not support that they are same construct. These results show that relationships present between the proposed model's antecedents and the construct perceived relationship rewards are not significant if perceived relationship rewards is replaced by long-term relationship intention. Figure 2 shows that all of the model's antecedents significantly affect perceived relationship rewards. This is not the case for the long-term relationship intention construct (see Figure 4). This lack of consistency in predictive validity supports the theoretical difference in the constructs. Additionally, deleting long-term relationship intention (option B) from the research model would leave the research model without any implications on actual human behavior. Deleting the intention construct from the model would greatly decrease the theoretical and practical contribution of the model, especially with the lack of support of predictive validity between the two constructs.



Fit statistics		SRMR	.055
χ^2	522.501	CFI	.964
S-B χ^2	327.739	RMSEA	.062 (.052,.071)

* Indicates paths not significant at $p<.05$

Figure 4. Structural Model: Standard Regression Weights, Variance Explained, and Fit Statistics for Option C

Cases like this may certainly not be very common in IS research. Additionally, this example shows how a theoretical understanding of the model can help a researcher to navigate testing the assumptions of possible different options for these rare circumstances (e.g., Table 12). However, this example is an effective means of demonstrating some of the strengths and weaknesses of the commonly used discriminant validity analysis techniques in IS research. It has also been helpful to show the importance of theory and the triangulation of data analysis techniques for interpretation purposes. Table 13 summarizes these results and shows that option B is the most defensible considering all the empirical results presented here. However, option D does also have empirical support and that of being theoretically justified.

Table 13. Results of Comparison Options for Dealing with Discriminant Validity Concerns between Long-Term Relationship Intention and Perceived Relationship Rewards

		Theoretical support	Empirical support
Option A	Combine the factors and test the model with all the manifest variables from both Perceived Relationship Rewards and Long-Term Relationship Intention to one single construct.	None	Limited
Option B	Delete Long-Term Relationship Intention from the model and have Perceived Relationship Rewards be the ultimate exogenous variable.	Limited	Supported
Option C	Delete Perceived Relationship Rewards from the model and have Long-Term Relationship Intention as the ultimate exogenous variable.	None	None
Option D	Leave the model as proposed and tentatively accept the high correlation between the factors as the estimated relationship between these two constructs.	Supported	Supported with χ^2 not AVE

5 Recommended Method for Comprehensive Theory-Driven Discriminant Validity Analysis

Considering the strengths and weaknesses of the commonly used methods of discriminant validity analysis that we outline above, we propose a method for analyzing discriminant validity for cross-sectional data in theory-driven research. This method stresses using multiple analysis techniques and sensitivity to the theoretical perspectives in interpreting results of discriminant validity analyses. We used multiple analysis techniques to lessen the impact of weaknesses of any one.

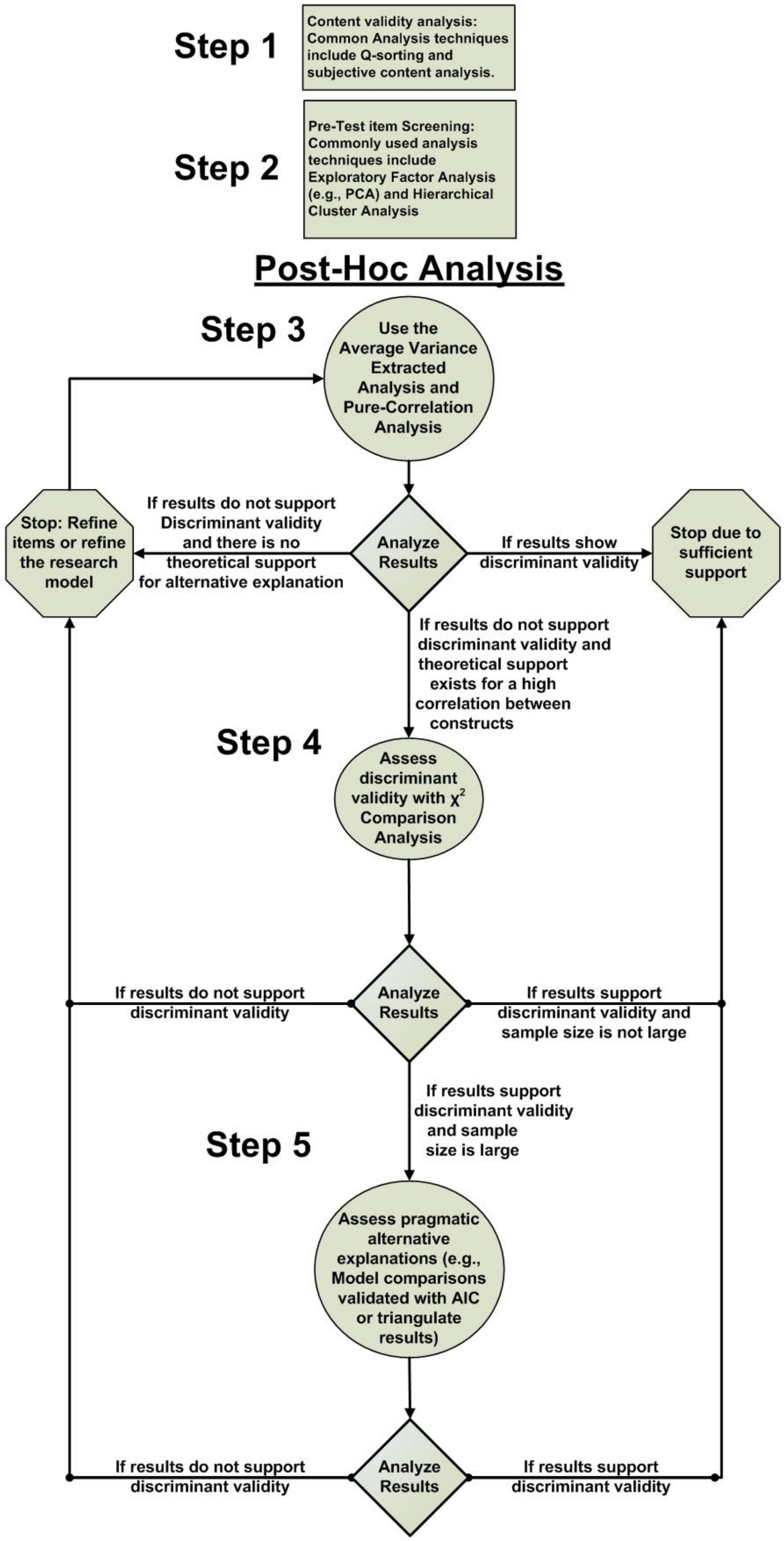


Figure 5. Decision Tree for Assessing Discriminant Validity with Cross-sectional Data

We need to acknowledge the limitations of this proposed method (see Figure 5). First, to complete this method, we assume analysis is being conducted in a context of covariance-based SEM. There are many justifiable alternatives in which analysis would not use covariance-based SEM analysis techniques. Therefore, this approach is not applicable in every research context. Second, this approach is focused on post-hoc analysis and, therefore, assumes that content analysis (step 1) and item screening (step 2) (e.g., Q-sorting or hierarchical cluster analysis) have already been conducted. With these boundaries in mind, we posit a theoretical understanding is necessary to appropriately use these analysis techniques and interpret results. Referring to Figure 5, we discuss the three steps proposed for post-hoc analysis.

Step 3: we first recommend using the AVE technique in conjunction with the correlation-based analysis to assess if any constructs are shown not to meet the generally accepted threshold. Analyzing the results from these techniques will determine if further analysis is needed. If both the correlation between the constructs and the AVE comparison support discriminant validity, then no further action is need; such results support discriminant validity. However, if discriminant validity is not supported, further action may be required to determine the viability of the items and possibly the theoretical perspective. If there is no theoretical justification for a high correlation between constructs, then the measurement items should be refined. If there is a theoretical justification for the relationship between two highly correlated constructs, then we recommend moving to step 4.

Step 4: interpreting the results from a χ^2 comparison needs to be tempered with respect to the sample size and level of deviation from a normal distribution. We propose three recommendations at this point, depending on the result of the analysis: a) If the results do not support discriminant validity, we recommend stopping to refine the measurement items; b) If this analysis supports discriminant validity, the sample size is not considered large, and there is no concern of non-normality in the data, we recommend stopping here. Limited and sufficient support can be offered in support of discriminant validity; and c) If sample size is considered to be large or a significant level of non-normality exists in the data and results support discriminant validity, we propose moving on to the final step in this method.

Step 5: the final step includes a series of model comparisons that encapsulate all possible alternatives for representing the constructs in question. This comparison should be theoretically driven and can be verified using the AIC fit statistic with SEM. If step 5 does not yield results that support discriminant validity, we propose that a further refinement of the measurement items or construct definition is needed. However, if results of this analysis support discriminant validity, then we propose that this should be sufficient in cases of constructs that theory supports to be highly correlated.

6 Discussion and Conclusions

Note that a limitation of this work is that we do not cover an exhaustive list of analysis techniques for reliability, all forms of validity, and discriminant validity. Other analysis techniques can be found and would demonstrate their own strengths and weaknesses. It is possible that there may be other complimentary analysis techniques more appropriate in certain situations then those highlighted in this work. Another limitation is that the topics discussed in this paper can be used for an opposing intent. Researchers may choose to use the arguments presented in this paper to capitalize on the weaknesses associated with one of these methods (the one that supports their own hypothesis) over others without theoretically justifying and triangulating results. This is obviously not the intent of our suggestions. We encourage authors to use these techniques responsibly.

Future research in this area may focus on other analysis techniques for other types of data (non-cross-sectional data) and other areas of concern for instrument development (e.g., reliability). Another focus for this line of research is using multiple analysis techniques in other aspects of data analysis (e.g., using SEM for an omnibus model and ANOVA to show simple effects and interactions). Future research could also consider other areas in which analysis triangulation can be used.

References

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Beverly Hills, CA: Sage.

Anderson, J. C., & Gerbing, D. W. (1998). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*, 411-423.

Anderson, J. C., Gerbing, D. W., & Hunter, J. E. (1987). On the assessment of unidimensional measurement: Internal and external consistency, and overall consistency criteria *Journal of Marketing Research*, *24(*4), 432-437.

Aronson, E., & Linder, D. (1965). Gain and loss of esteem as determinants of interpersonal attractiveness. *Journal of Experimental Social Psychology*, *1*, 156-172.

Bagozzi, R. P. (2011). Measurement and meaning in information systems and organizational research: Methodological and philosophical foundations. *MIS Quarterl*y, *35(*2), 261-292.

Baker, B. (2000). *Some useful statistical tables*. Retrieved from http://duke.usask.ca/~rbaker/Tables.html

Bentler, P.M. (2005). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.

Bollen, K. A. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly*, *35(*2), 359-372.

Brown, S. R. (1986). Q technique and method. In W. D. Berry & M. S. Lewis-Beck (Eds.), *New tools for social scientists: Advances and applications in research methods* (pp. 57-76). Beverly Hills, CA: Sage.

Burns, G. L., Boe, B., Walsh, J. A., Sommers-Flanagan, R., & Teegarden, L. A. (2001). A confirmatory factor analysis on the DSM-IV ADHD and ODD symptoms: What is the best model for the organization of these symptoms? *Journal of Abnormal Child Psychology*, *29(*4), 339-349.

Burt, C., & Williams, E. L. (1962). The influence of motivation on the results of intelligence tests. *The British Journal of Statistical Psychology*, *15*, 129-136.

Byrne, B. M. (2006). Structural equation modeling with EQS: Basic concepts, applications, and programming. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, *13*, 287-321.

Campbell, D. E., Wells, J. D., & Valacich, J. S. (2009). Diagnosing and managing online business-to-consumer (B2C) relationships: Toward an ecommerce B2C relationship stage theory. *AIS Transactions on Human-Computer Interaction*, *1(*4), 108-132.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Carlson, D. S., Kacmar, K. M., & Williams, L. J. (2000). Construction and initial validation of a multidimensional measure of work-family conflict. *Journal of Vocational Behavior*, *56*, 249-276.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, *12,* 471-492.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295-336). Mahwah, NJ: Lawrence Erlbaum Associates.

Christie, A., Jordan, P., Troth, A., & Lawrence, S. (2007). Testing the links between emotional intelligence and motivation. *Journal of Management & Organization*, *13(*3), 212-226.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16-29.

Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *MIS Quarterly*, *35(*2), 335-358.

Everitt, B. S. (1993). Cluster analysis. London: Arnold.

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

Gefen, D., & Straub, D. (2005). A practical guide to factorial validity using PLS-graph: Tutorial and annotated example. *Communications of the Association for Information Systems*, *16*, 91-109.

Gefen, D., Straub, D. W., & Boudreau, M. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems*, *4,* 1-70.

Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.

Goleman, D. (1998). *Working with emotional intelligence*. New York: Bantam Books.

Gordon, A. D. (1999). *Classificatio*n. London: Chapman & Hall.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). Multivariate data analysis with readings. Englewood Cliffs, NJ: Prentice Hall.

Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structural analysis be trusted? *Psychological Bulletin*, *112*, 351-362.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Hughes, M. A., Price, R. L., & Marrs, D. W. (1986). Linking theory construction and theory testing: Models with multiple indicators of latent variables. *Academy of Management Review*, *11(*1), 129-144.

Huston, T. L., & Levinger, G. (1978). Interpersonal attraction and relationships. *Annual Review of Psychology*, *29*, 115-156.

Kline, R. B. (2005). Principles and practice of structural equation modeling. New York, NY: The Guilford Press.

Kloot, W. A. v. d., Spaans, A. M. J., & Heiser, W. J. (2005). Instability of hierarchical cluster analysis due to input order of the data: The permucluster solution. *Psychological Methods*, *10(*4), 468-476.

Levinger, G. (1980). Toward the analysis of close relationships. *Journal of Experimental Social Psychology*, *16(*6), 510-544.

Lewis, B. R., Templeton, G. F., & Byrd, T. A. (2005). A methodology for construct development in MIS research. *European Journal of Information Systems*, *14*, 388-400.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*, 53-76.

Loehlin, J. C. (2004). Latent variable models: An introduction to factor, path, and structural equation analysis. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Loiacono, E.T., Watson, R.T., & Goodhue, D.L. (2007). WebQual: An instrument for consumer evaluation of web sites. International Journal of Electronic Commerce, *11*(3), 51-87.

Loretan, J. (1965). The decline and fall of group intelligence testing. *Teachers College Record*, *67*, 10-17.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35(*2), 293-334.

Marsh, H. W., Hey, J., & Roche, L. A. (1997). Structure of physical self-concept: Elite athletes and physical education students. *Journal of Educational Psychology*, *89*, 369-380.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Implications for educators* (pp. 3-31). New York: Basic Books.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13(*3), 334-359.

Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher*, *10(*9), 9-20.

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, *2(*3), 192-222.

Moss, J. (2005). Race effects on the employee assessing political leadership: A review of Christie and Geis' (1970) Mach IV measure of Machiavellianism. *Journal of Leadership & Organizational Studies*, *11(*2), 26-33.

Muir, M. S., & de Charms, R. (1975). Personal causation training and goal setting in seventh grade. In *Proceedings of the American Educational Research Association.*

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, *31(*4), 623-656.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903.

Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, *20*, 134-148.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111-163.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for Chi-square statistics in covariance structure analysis. In *Proceedings of the American Statistical Association 1988 Proceeding of the Business and Economic Sections.*

Segars, A. (1997). Assessing the unidimensionality of measurement: A paradigm and illustration within the context of information systems research. *Omega*, *25(*1), 107-121.

Segars, A. H., & Grover, V. (1998). Strategic information systems planning success: An investigation of the construct and its measurement. *MIS Quarterly*, *22(*2), 139-163.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin Company.

Storey, V., Straub, D., Stewart, K., & Welke, R. (2002). A conceptual investigation of the electronic commerce industry. *Communications of the ACM*, *43(7*), 117-123.

Straub, D., Boudreau, M., & Gefen, D. (2004). Validation guidelines for is positivist research. *Communications of the AIS*, *13(*24), 380-427.

Straub, D. W. (1989). Validating instruments in MIS research. *MIS Quarterly*, *13(*2), 147-169.

Straub, D. W., & Burton-Jones, A. (2007). Veni, vidi, vici: Breaking the TAM logjam. *Journal of the Association for Information Systems*, *8(*4), 223-229.

Thomas, D. M., & Watson, R. T. (2002). Q-sorting and MIS research: A primer. *Communications of the Association for Information Systems*, *8(*9), 141-156.

Vallerand, R. J., & Richer, F. (1988). On the use of the causal dimension scale in a field setting: A test with confirmatory factor analysis in success and failure situations. *Journal of Personality and Social Psychology*, *54(*4), 704-712.

Vandenberg, R., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3(*1), 4-69.

Vogt, W. P. (1999). Dictionary of statistics & methodology: A nontechnical guide for the social sciences. Thousand Oaks, CA: Sage.

Werts, C. E., Linn, R. L., & Joreskog, K. (1974). Interclass reliability estimates: Testing structural assumptions. *Educational and Psychological Measurement*, *34*, 25-33.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In ,R. H. Hoyle (Ed.) *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.

Appendix A: Items Used in Survey Instrument for Survey 2

Table A1. List of Items

Measure name	Code	Item
Long-term relationship intention	C1_1	I would like to be in a long-term customer relationship with this company.
	C1_2	Engaging in a long-term relationship with this organization would prove beneficial to me as a consumer.
	C1_3	Assuming that I was interested in one of their products, I could see myself initiating a long-term relationship with this organization.
Perceived relationship rewards	C2_1	Doing business with this organization would be a rewarding experience.
	C2_2	Customers most likely find doing business with this organization to be a rewarding experience.
	C2_3	I feel that there are more positive consequences than negative in dealing with this company.
Visual appeal (Loiacono et al., 2007)	C3_1	The website is visually pleasing.
	C3_2	The website displays visually pleasing design.
	C3_3	The website is visually appealing.
Competent behavior (*reverse coded)*	C4_1	This website does not function competently.
	C4_2	This website is not adequate in doing what it is supposed to do.
	C4_3	This website doesn't do what it is supposed to do.
	C4_4	This website does not behave in a competent manor.
	C4_5	This website is not a competent interface for representing and selling this organization's products.
Relationship compatibility	C5_1	Based on this website, I believe that this company's beliefs and values are similar to mine.
	C5_2	Based on this website, I believe that this organization and I have harmonious beliefs and values.
	C5_3	I agree with this company's beliefs.
	C5_4	I agree with this company's values.
	C5_5	My perspective on ethics and values seems to be aligned with those of this organization.
Relationship receptiveness	C6_1	It was very obvious that this company really wanted me as a customer.
	C6_2	Based on this website, I think that this company is trying to get as many customers as it possibly can, and would like me to be a long-term customer.
	C6_3	This firm really desires me to be their customer.
	C6_4	Based on this website, I think that this company really wanted me to be a long-term customer.

Appendix B: Exploratory Factor Analysis Results

Table B1. Final Loading Matrix for the Four Antecedent Variables

	Factor			
	1	**2**	**3**	**4**
C3_12	**.885**	.162	.246	.180
C3_10	**.868**	.222	.239	.226
C3_9	**.861**	.167	.258	.197
C3_8	**.860**	.174	.316	.223
C3_7	**.856**	.194	.197	.250
C3_11	**.855**	.205	.240	.226
C3_5	**.842**	.192	.307	.237
C3_14	**.822**	.208	.258	.178
C3_3	**.810**	.193	.288	.240
C5_14	.168	**.846**	.186	.283
C5_16	.188	**.832**	.199	.245
C5_1	.186	**.830**	.273	.194
C5_4	.233	**.828**	.249	.196
C5_11	.174	**.816**	.208	.257
C5_17	.198	**.815**	.267	.250
C5_15	.167	**.791**	.218	.287
C5_10	.183	**.770**	.138	.208
C4_12	.260	.240	**.829**	.295
C4_7	.301	.259	**.790**	.265
C4_11	.333	.270	**.768**	.230
C4_10	.280	.266	**.721**	.329
C4_6	.260	.264	**.714**	.342
C4_8	.294	.266	**.704**	.342
C4_16	.281	.240	**.704**	.267
C4_9	.331	.200	**.698**	.215
C4_5	.297	.175	**.688**	.341
C6_6	.219	.266	.322	**.758**
C6_15	.259	.366	.306	**.730**
C6_9	.289	.353	.254	**.725**
C6_8	.278	.322	.357	**.690**
C6_12	.256	.288	.221	**.674**
C6_14	.143	.254	.302	**.668**
C6_5	.353	.343	.383	**.650**
C6_7	.313	.199	.434	**.635**
C6_13	.333	.310	.373	**.635**
Extraction method: maximum likelihood. Rotation method: varimax with Kaiser normalization.				

Table B2. EFA for the Mediating Variable

	Factor
	1
C2_1	.705
C2_10	.776
C2_11	.657
C2_12	.904
C2_13	.799
C2_14	.827
C2_15	.759
C2_2	.795
C2_3	.753
C2_4	.900
C2_5	.822
C2_7	.870
C2_8	.842
C2_9	.507
Extraction method: maximum likelihood.	

Table B3. EFA for the Dependent Variable

	Factor	
	1	2
C1_15	.807	.312
C1_13	.790	.326
C1N_15	.782	.354
C1_17	.777	.466
C1_8	.738	.500
C1_16	.734	.500
C1_7	.697	.560
C1N_13	.676	.488
C1N_2	.652	.453
C1_5	.650	.546
C1N_4	.629	.461
C1_14	.599	.559
C1N_14	.563	.509
C1N_3	.531	.378
C1N_5	.517	.364
C1_12	.389	.746
C1_3	.471	.699
C1_11	.382	.675
C1_9	.583	.647
C1_2	.437	.636
C1_4	.622	.629
C1N_6	.216	.614

Table B3. EFA for the Dependent Variable

C1N_9	.429	.612
C1_6	.608	.612
C1N_11	.552	.611
C1_1	.354	.603
C1N_8	.453	.585
C1_10	.230	.547
C1N_10	.431	.545
C1N_12	.427	.537
C1N_7	.437	.488
C1N_1	.286	.358
C1N_17	.329	.357
C1N_16	.170	.238
Extraction method: maximum likelihood. Rotation method: varimax with Kaiser normalization.		

Appendix C: AVE and Composite Reliability Calculation Example for Microsoft Excel

Figure C1 shows an example of how composite reliabilities and AVE can be conducted using Excel. First, standardized loadings and correlations should be imported (copy/pasted) from an SEM output. Then, the using the formulas like those given in Figure C1, AVE can be calculated. Comparing those AVE values with the squared correlations gives the comparison of AVE to squared correlations commonly found in IS research.

	A	B	C	D
1	Standardized Loadings		Composite Reliabilities	AVEs
2	Cbhn12	0.782	0.871	0.693
3	Cbhn16	0.901		
4	Cbhn5	0.809		
	Bi1	0.949	0.976	0.932
	Bi2	0.974		
	Bi3	0.973		
	Relcp10	0.907	0.968	0.884
	Relcp15	0.924		
	Relcp16	0.961		
	Relcp17	0.968		
	Relcp4	0.94		
	Rlrep12	0.869	0.957	0.847
	Rlrep14	0.907		
	Rlrep15	0.956		
	Rlrep8	0.947		
	Prewp2	0.927	0.948	0.858
	Prewp5	0.931		
	Prewp7	0.921		
	VAP1	0.973	0.984	0.953
	VAP2	0.973		
	VAP3	0.982		

Table 6. EQS Estimated Squared Correlations and (AVE)*

	CB	VAP	RLCMP	RLRSP	PR	BI
CB	**0.69**					
VAP	0.15	**0.95**				
RLCMP	0.12	0.38	**0.88**			
RLRSP	0.11	0.45	0.53	**0.85**		
PR	0.21	0.54	0.77	0.56	**0.86**	
BI	0.14	0.46	0.58	0.46	0.73	**0.93**

* AVE figures are shown in bold along the diagonal

Correlations

	CB	VAP	RLCMP	RLRSP	PR
VAP	0.392				
RLCMP	0.341	0.613			
RLRSP	0.333	0.67	0.728		
PR	0.462	0.738	0.876	0.747	
BI	0.369	0.678	0.761	0.678	0.852

Formula in cell 'C2' calculating composite reliabilities:

(B2+B3+B4)*(B2+B3+B4)/((B2+B3+B4)*(B2+B3+B4)+((1-(B2*B2)+(1-(B3*B3))+(1-(B4*B4)))))

Formula in cell 'D2' calculating AVE:

((B2*B2)+(B3*B3)+(B4*B4))/(((B2*B2)+(B3*B3)+(B4*B4))+((1-(B2*B2))+(1-(B3*B3))+(1-(B4*B4))))

Figure C1. Example of AVE and Composite Reliability Calculations in Excel

Appendix D: Conducting a χ² Comparison Analysis

χ^2 comparison analysis compares the χ^2 values between fixed and free solutions for each pair of constructs being assessed in a measurement model using structural equation modeling (SEM) analysis (Anderson & Gerbing, 1988a; Segars, 1997; Straub et al., 2004). The free solution is a measurement model that has no fixed parameters in the correlations between the factors (see Figure D1, emphasis added). The fixed solutions are those that fix one correlation between two factors to one. All possible fixed solutions must be evaluated. Then, a $\Delta\chi^2$ analysis is conducted comparing the free solution to all the fixed solutions. To show discriminant validity, the free solution should show a significant improvement in fit. This would show that modeling the constructs separately is a better fit to the data than considering them the same, meaning that all possible combinations of factors indicate a significantly worse fit to the data than separating the constructs (Anderson & Gerbing, 1988a; Segars, 1997; Straub et al., 2004).

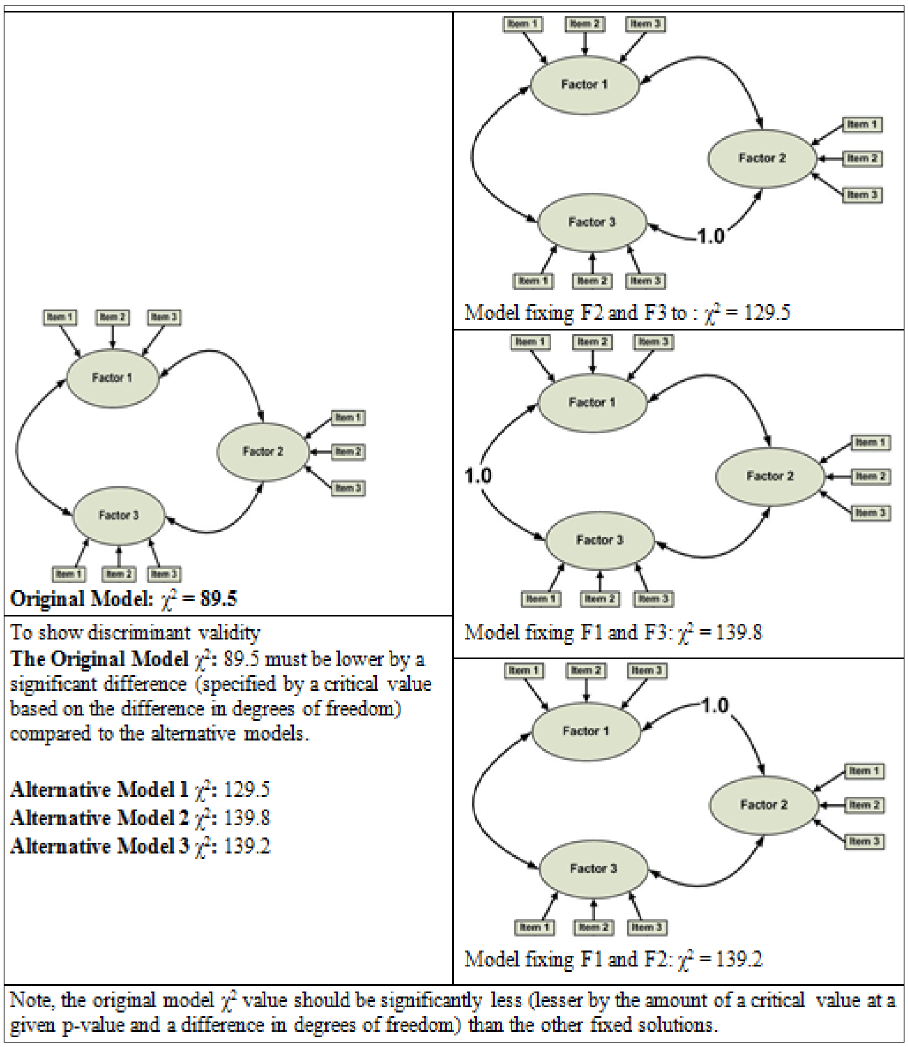


Figure D1. Example of a χ^2 Comparison Analysis for Discriminant Validity

To make comparisons across the fixed and free solutions, one needs to derive the critical value. Critical values are usually obtained using a look-up table for χ^2 critical values that are common in statistics books and online (e.g., Baker, 2000). To look up this critical value, one needs to specify the degrees of freedom (df) and the p-value desired (e.g., .05 or .001). For this type of test, the difference between the original model (free solution) and any of the fixed models will be 1 df. Therefore, researchers only need to determine the desired p-value. By comparing the χ^2 values between the original model (free solution) and all fixed solutions, the original model should have a χ^2 value that is significantly less than any of the fixed solutions. For example, Figure D1 shows a χ^2 comparison for a model with three constructs. χ^2 difference test for 1 df at the p<.001 requires a difference of 10.828 in the χ^2 values (Baker, 2000). Therefore, this χ^2

comparison analysis supports discriminant validity because the original model that shows three distinct factors fits the data significantly better than any alternative model combining factors (fixing correlations).

Another form of this type of discriminant validity test would be to fix the correlations at zero. To fix the correlations at one conceptually means that the constructs are identical or have a correlation of one. Fixing the correlations at zero offers the conception that the constructs are maximally different and have a correlation of zero. If this test fixed the correlations to zero, one would not want a significant decrease in fit. In other words, when analyzing the results, the researcher would want to show that fixing the correlations at zero did not cause a significant decrease in fit for discriminant validity to be upheld.

If a researcher is data driven and trying to identify factors from the data, it may be more conservative to use the test that fixes the correlations at zero. Fixing the correlations to zero may also be useful in situations where constructs are not theorized to be highly correlated. Fixing the correlations to zero may also be appropriate in theory-driven research if the correlations between constructs are not expected to be significantly different than zero. However, if the researcher is conducting a theory-driven analysis and expects that constructs would be significantly correlated yet still distinct constructs, then fixing the correlation to one would be the analysis recommended here because one would not expect the correlations to be zero (constructs to be maximally different).

About the Authors

Damon Campbell is an Associate Professor and Kelly Gene Cook, Sr. Chair of Business Administration in the Else School of Management at Millsaps College. He holds a PhD and MBA from Washington State University. His primary research interests include eCommerce, human computer interaction, and interface design. His research has appeared in *Information Systems Research*, *Decision Sciences*, *Journal of the Association for Information Systems*, and other journals and conferences.

John Wells is an Associate Professor in the Isenberg School of Management at the University of Massachusetts Amherst. He received his BBA in management from the University of Oklahoma and his MS and PhD in management information systems from Texas A&M University. He has worked as a systems engineer for Electronic Data Systems and the Oklahoma State Senate. His active research areas are e-commerce strategy and interface design. His work has appeared in such journals as *Information Systems Research*, *Decision Sciences*, *Journal of Management Information Systems*, *European Journal of Information Systems*, and *Information & Management* as well as in a number of international conferences.

Craig Parks is Professor of Psychology at Washington State University, where he teaches introductory and advanced analysis and methodology. He received his PhD in Psychology from the University of Illinois, with an emphasis in social psychology and minor in quantitative analysis. He has been the editor of *Group Dynamics*.

JITTA

JOURNAL OF INFORMATION TECHNOLOGY THEORY AND APPLICATION

JITTA is a Publication of the Association for Information Systems
ISSN: 1532-3416