Communications of the Association for Information Systems

Volume 12 Article 2

July 2003

Assessing Unidimensionality Through LISREL: An Explanation and an Example

David Gefen

Drexel University, gefend@drexel.edu

Communications of the Association for Information Systems

ASSESSING UNIDIMENSIONALITY THROUGH LISREL: AN EXPLANATION AND EXAMPLE

DAVID GEFEN
LeBow College of Business
Drexel University
gefend@drexel.edu

ABSTRACT

Research in MIS often focuses on the relationships among latent variables of interest that cannot be directly measured. Because of potential error in measurement and associated confounding, indirect measurement of latent constructs requires formal assessments of reliability and validity. Without these measures, resultant paths in causal implications may be inaccurate, biased, and unstable. However, even with favorable metrics of validity and reliability, it is still possible for estimated models to be confounded. In many cases, such confounding occurs when a measurement item reflects more than one latent construct, that is, when there is a lack of unidimensionality. This problem can lead to false assumptions regarding the strength of paths between latent constructs and patterns of causality within a nomological network. While assessing unidimensionality is a critically important aspect of validity, it is not always formally tested in MIS research.

This tutorial introduces the concept of unidimensionality from a LISREL Confirmatory Factor Analysis (CFA) perspective. Assuming that the underlying data distribution assumptions and model used are correct, the tutorial provides a step-by-step example of how to assess unidimensionality with LISREL. The tutorial also shows how a CFA can detect problematic multidimensional items and the problems that can occur if undetected.

Keywords: Research methodology, LISREL, SEM, unidimensionality, reliability, validity, e-commerce, gender studies.

I. INTRODUCTION

Many constructs, such as those dealing with perceptions and beliefs, cannot be measured directly. These constructs are often approximated as *scales* in linear regression models and, if correctly modeled, more appropriately estimated as *latent variables* in Structural

Equation Modeling.[1] In lieu of structural equation modeling (SEM) estimation, the standard procedure is to represent the latent variable as a mean (in the case of linear regressions) or as a weighted mean (based on SEM) of several items that can be measured directly. These directly measurable items are called *measurement items*, which are also known as *indicators* or *item measures*.

Instead of modeling latent variables as a function of measured items, SEM treats measured items as functions of latent variables. Accordingly, it is essential to verify that the measurement items reflecting each latent variable are

(1) consistent with each other (scale reliability),

(2) reflect the same latent variable of interest (convergent validity) while

(3) making it statistically distinct from other latent variables (discriminant validity), and that

(4) the variance each item shares with other items does not relate to an unspecified latent variable (unidimensionality).

When there is inadequate reliability, insufficient convergent, and discriminant validity, or a notable lack of unidimensionality, the conclusions drawn from the statistical analysis may be unwarranted, biased, and unstable [Gefen et al., 2000, Gerbing and Anderson, 1988, Nunnally and Bernstein, 1994, Segars, 1997].

While researchers in MIS are generally aware of the need to establish reliability, convergent validity, and discriminant validity, very few studies establish properties of unidimensionality for item measures.[2] For example, a search of ABI-INFORM for the years 1999 through 2001 showed that only about a dozen articles in MIS journals that include the term unidimensionality and that only about two dozen more contained the term between 1986 and 1999 and that almost all of these are limited to the top MIS journals.[3] In all these studies unidimensionality was assessed only through a CFA with LISREL.

Heeding the mostly unanswered calls of Al Segars [1994, 1997] some years back to include unidimensionality testing in MIS research, the objective of this tutorial is to provide a didactic introduction with appropriate examples highlighting the importance of examining unidimensionality in determining measurement efficacy. The examination of unidimensionality is demonstrated with LISREL.[4] To this end, the tutorial first explores the concept of unidimensionality, contrasts it with traditional measures of reliability and validity and then demonstrates the usefulness of LISREL for inferring its existence. Assuming that the underlying assumptions of LISREL are correct – namely that there is a multivariate normal distribution,[5] that the measures are continuous,[6] that the relationships between each measurement item and its

---

[1] For a detailed discussion of SEM, its terminology, and a comparison with linear regression see Gefen et al. [2000] in CAIS.
[2] For a detailed discussion on current MIS research practices see Gefen et al. [2000] and Boudreou et al. [2001].
[3] MISQ takes the lead here with 14 articles over the 15 year period with Decision Sciences at 11 and JMIS at 9 and The Database for Advances in Information Systems and Omega with 2 each.
[4] This tutorial deals with unidimensionality from a LISREL CFA perspective. Unidimensionality is also discussed in the context of Item Response Theory (IRT), see discussion in Hatti [1985] and in Hambleton et al. [Hambleton et al., 1984, Hambleton et al., 1991].
[5] Because of its the distribution assumptions, dichotomous data are problematic in LISREL and require a tetrachoric correlation matrix, rather than a Pearson correlation matrix [Jöreskog and Sörbom, 1993, Schumacker and Beyerlein, 2000].
[6] Although technically the underlying assumption in LISREL is that the measurement items are continuous, it is an accepted practice to allow Likert type scales too.

latent construct is linear, and that the measurement items exhibit local independence[7] – the technique can provide a clearer "line of sight" between the property of unidimensionality and the compliance or deviation of item measures. It should be noted that when there is a serious deviation from these assumptions, the conclusions drawn, as in other LISREL analyses may be biased [Jöreskog and Sörbom, 1989]. The context for our empirical examples is the Gefen and Straub [1997] model of gender and IT adoption, examined here in the context of purchasing flight tickets online.

II. THE THEORY

Unidimensionality can be defined as a concept and through mathematics. From a conceptual viewpoint, consider an example of an IQ test that measures two dimensions of intelligence: quantitative and verbal. Both dimensions of IQ cannot be measured directly in the way temperature or pressure can be measured. Instead, IQ is measured through a battery of indirect measures that reflect its various dimensions. Since each such measure only approximates the actual intelligence type is reflects, it is necessary to measure each dimension of intelligence through several items to capture its variance accurately and in an unbiased manner [Gefen et al., 2000, Gerbing and Anderson, 1988, Segars, 1997].

In this example, these two dimensions of IQ are measured through 8 items, four for each of the dimensions. Each measurement item is a numeric score in one examination with a mark ranging from 0 to 100. Table 1 contains the measurement items. The latent variable is calculated as the mean or the weighted mean of the examinations that reflect it.

Table 1. Hypothetical Measurement Items in the IQ Test

QUANTITATIVE IQ
Item 1: Score in integral mathematics
Item 2: Score in algebra
Item 3: Score in trigonometry
Item 4: Score in mathematical problem solving of story quizzes
VERBAL IQ
Item 5: Score in English grammar
Item 6: Score in essay composition
Item 7: Score in reading comprehension
Item 8: Score in vocabulary

In an ideal test, the variance of each item would reflect only the variance of the construct it represents, that is, the score in English grammar would reflect only the student's English grammar aptitude. Of course, that is never the case because each measure introduces an element of measurement error into its variance – additional variance that is not related to the underlying latent construct, in this case, verbal intelligence. For example, if the room temperature was stifling hot when the examination was taken, then the excessive heat might also have contributed to the examination score and thus influenced the variance of the measurement item.

[7] Local independence means that the measurement items of a given latent variable are statistically independent of each other except for the variance that is related to the latent construct [Hatti, 1985, Hatti et al., 1996, Lord and Novick, 1968, McDonald, 1981].

In a more detailed representation, the variance of each of the measurement items can be described graphically as containing a combination of:

(1) *common variance*,
(2) *specific variance,* and
(3) *shared error variance.*

The non-common variance (i.e., specific and shared error variance) is also known as *residual variance* in covariance-based SEM. Common variance is variance that the measurement item shares with its underlying construct and that theoretically reflects the latent variable variance. As a rule of thumb, the common variance should be at least 50% [Hair et al., 1998]. In the example shown in Figure 1, common variance with the first latent construct, quantitative IQ, is marked in red, and common variance with the second latent construct, verbal IQ, is marked in Blue. The second component of measurement item variance is its specific variance. This variance is the unique error variance relating of the specific measurement item, hence its name. This variance is not shared with other measurement items and appears as yellow in Figure 1.

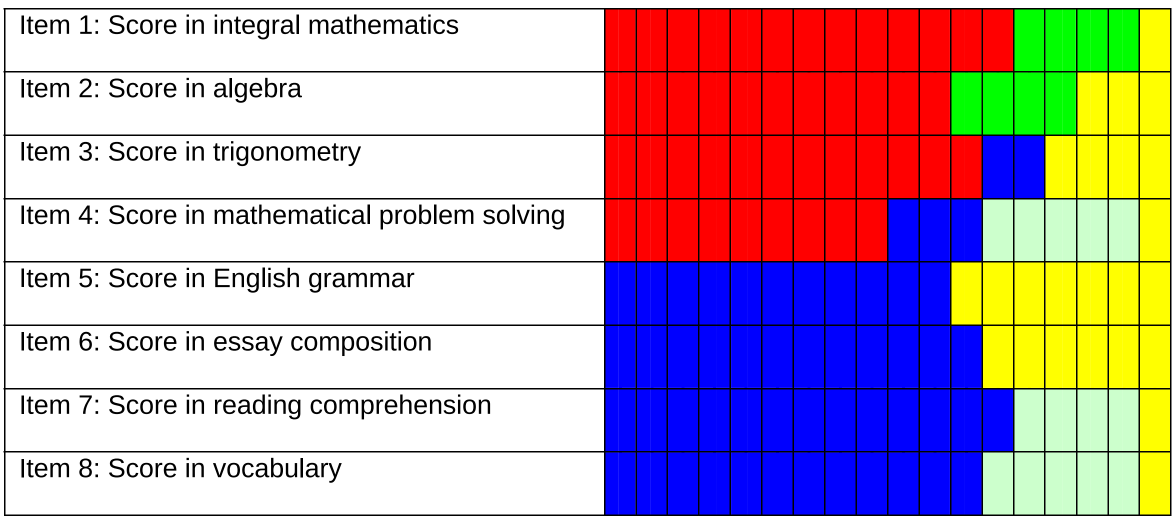


Figure 1. Variance of Measurement of Hypothetical IQ Items by Type

Shared error variance, which is non-common variance that is shared with other measurement items, is marked in hues of green. This error variance may occur when a common source of variance affects more than one measurement item, either because of a latent variable that is not accounted for in the model or because of some other factor affecting the measurement error in more than one measurement item, such as unfavorable examination conditions. When such non-common variance is significantly correlated with the non-common variance of other measurement items, then the items may not exhibit unidimensionality. Further, resultant models may not represent the truest effects among latent variables [Segars, 1997]. Structural Equation Modeling is a useful tool for discovering these potential threats. Such a case can be seen graphically in the shared variance between items 4, 7, and 8 in Figure 1, which may represent an unaccounted for latent variable such as reading comprehension. Shared variance between items 1 and 2 may represent another unaccounted for latent variable such as number-crunching aptitude. Such shared variance violates unidimensionality [Hatti, 1985]. Also note that some of the variance in the quantitative scores is related to verbal intelligence. This observation seems reasonable because reading comprehension should affect these scores.

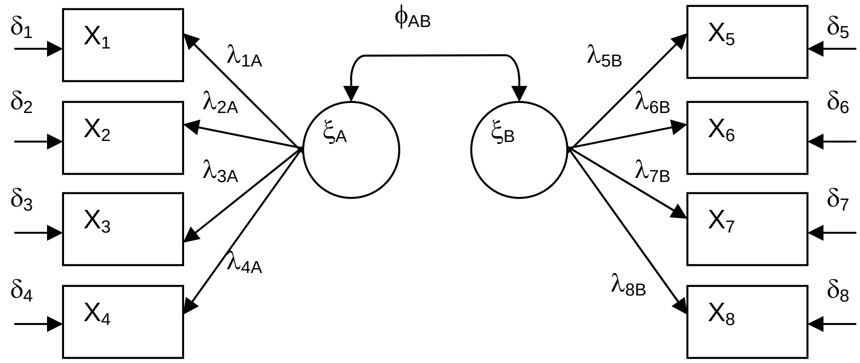


Figure 2. SEM Model of Hypothetical IQ Items

As illustrated in Figure 2, the measurement items, $X_{1..8}$, are combined to form the latent variables, $\xi_{1..2}$, in accordance with the theoretical research model. Measurement errors, $\delta_{1..8}$, are not significantly correlated with each other. All item loadings, $\lambda_{1..8}$, reflect only their respective latent variables. This graphical representation illustrates the operationalization of theory and the essential ingredients of sound measurement. We now discuss these properties from the perspective of their underlying mathematics.

UNIDIMENSIONAL VALIDITY

Behind every measurement item there should be one and only one underlying construct [Gorden, 1977, Hatti, 1985, Hatti et al., 1996]. In other words, each measurement item should reflect only its associated latent construct without significantly reflecting any other construct. Accordingly, the non-common variance of each measurement item should ideally be the only measurement error and should not be significantly correlated with the non-common variance of any other measurement item.[8] This property is called unidimensionality [Gerbing and Anderson, 1988, Hatti, 1985]. For example, if δ_1 and δ_4 in Figure 2 were significantly correlated, then the unidimensionality of X_1 and X_4 would be suspect and the scale quality of ξ_A problematic.[9] In that case a possible third construct responsible for the significant residual variance could not readily be ruled out.

Related to this property is the local independence of the measurement items. Local independence is held when all the measurement items that reflect a given trait ξ are statistically independent of each other, that is the only shared factor underlying all these measurement items is ξ and all the remaining variance in the measurement items is random noise [Lord and Novick, 1968]. Consequently, this principle requires that the covariance between any pair of measurement items be zero and that all the higher statistical moments be products solely of the univariate moments, i.e., not of any combination among the moments of different measurement items [Hatti et al., 1996]. This situation is the case whether the relationships between the measurements items and ξ are linear or nonlinear [Hatti et al., 1996]. A slightly less demanding definition of local independence suggested by McDonald [1979], and known as the “weak principle of local independence” [Hatti et al., 1996, p. 2], is that only the covariance among these measurement items should be zero. The latter is assumed in LISREL CFA [Jöreskog and

[8] The “standard” threshold being a standardized residual above 2.58, which is a p-value < .01 [Jöreskog and Sörbom, 1989].

[9] The error variance of the X measurement items is labeled δ, called theta delta. The error variance of the Y measurement items is labeled ε, called theta epsilon.

Sörbom, 1989] and is the definition applied in this tutorial. Mathematically, the weak principle of local independence requires that there should be a latent variable ξ such that on average the conditional covariance of all measurement item pairs is close to zero [Stout, 1987]. Unidimensionality is the operationalization of this weak principle of local independence [Hatti et al., 1996] with Stout's T [Stout, 1987] being a non-LISREL index of this measure of weak local independence. (See Hatti et al. [1996] for a discussion of this technique and criticism of it.)

In Anderson and Gerbing's [1988] formulation of unidimensionality, if each set of measurement items has only one underlying construct, ξ, then, assuming a linear relationship, each measurement item X_i is given by the product of its factor loading, λ_i, on its latent variable combined with its residual, δ_i, which is assumed to be have no significant correlation with any other X or δ [Gerbing and Anderson, 1988, Jöreskog and Sörbom, 1989, Segars, 1997]. In other words, each measurement item is assumed to be a linear reflection of its latent construct combined with random error:i

$$X_i = \lambda_i\xi + \delta_I \quad (1)$$

Assuming linear relationships, a measurement item X_i will be unidimensional if two rules apply.[10]

1. It must show internal consistency with other measurement items of the same latent variable. Internal consistency is shown when its correlation with any other measurement item, X_j, of the same latent variable, depict as ρ_{ij}, is equal to the product of the correlation of each measurement item with the latent variable, depict as $\rho_{i\xi}$ and $\rho_{j\xi}$:

$$\rho_{ij} = \rho_{i\xi} * \rho_{j\xi} \quad (2)$$

That is, the correlation between the two measurement items depends only upon their correlation with the latent variable, in effect meaning that the non-common variance of the measurement items, δ_i and δ_j, do not contribute significantly to the correlation between the two measurement items.

2. For X_i to be unidimensional it must show external consistency. External consistency is shown when its correlation with a measurement item, X_p, of any other latent variable, ξ_p, is equal to the product of the correlation of each measurement item with its latent variable, $\rho_{i\xi}$ and $\rho_{j\xi p}$, multiplied by the correlation of the two latent constructs, $\rho_{\xi\xi p}$:

$$\rho_{ip} = \rho_{i\xi} * \rho_{\xi\xi p} * \rho_{j\xi p} \quad (3)$$

That is, the correlation between the two measurement items that reflect different latent variables should depend only on their correlation with their respective latent variables and the correlation between the two latent variables. Again, in effect, meaning that the non-common variance of the measurement items, δ_i and δ_p, do not contribute to the correlation between the two measurement items. Given these two rules, it is possible to see that traditional metrics of measurement efficacy (reliability, convergent validity, discriminant validity) cannot assess unidimensionality, and why, consequently, unidimensionality should be examined separately [Gerbing and Anderson, 1988, Jöreskog and Sörbom, 1989, Segars, 1997]. We illustrate this point in the following sections.

RELIABILITY

Reliability measures the *internal consistency* of a latent variable, the degree to which several measurement items that reflect it are inter-correlated [Hair et al., 1998, Nunnally and Bernstein, 1994]. In essence, reliability measures the degree that the measurement items that reflect the same latent variable are in agreement with one another [Campbell and Fiske, 1959, Churchill,

---

[10] Linear relationships, linearity, is the underlying assumption behind covariance-based SEM as well as linear regression and ANOVA models [Hair et al., 1998].

1979]. The most commonly used measure of reliability in non-SEM analyses is Cronbach's α [1951], which is the de-facto measure of scale reliability [Peterson, 1994]. According to Churchill's seminal work, "Coefficient alpha *absolutely* should be the first measure one calculates to assess the quality of the instrument" [Churchill, 1979, p. 68 (italics in the original)]. When the reliability of a latent variable is low the standard practice is to drop items until the coefficient reaches the desired threshold [Churchill, 1979]. Coefficient α measures the average ratio of item variance to scale variance, accounting for the number of items in the scale [Cronbach, 1951]:

$$\alpha = \frac{k}{k-1} * (1-\Sigma\alpha_i^2/\alpha_s^2) \quad (4)$$

Where "k" is the number of items in the scale, α_i^2 the variance of item "i", and α_s^2 the variance of the scale. The accepted threshold for Cronbach's α is .80 [Nunnally and Bernstein, 1994], although even lower values (in the .60s) are common [Peterson, 1994]. Coefficient α is thus an estimate of the ratio of true latent variable variance to its observed variance [Hatti, 1985].

An alternative to coefficient α is *composite factor reliability* which is also an estimate of the ratio of common variance, $(\Sigma\lambda_i)^2$, to the total variance, $(\Sigma\lambda_i)^2 + \Sigma\delta_i$. Composite factor reliability does not assume equal loading of each measurement item on the latent variable, as Cronbach's α does. Rather, it evaluates the relative weight of each measurement items according to its estimated loading on the latent variable given the overall measurement model. Because of this, composite factor reliability often gives higher estimates of reliability than Cronbach's α does. Composite factor reliability is calculated using the following equation [Werts et al., 1974]:

$$\rho = \frac{(\Sigma\lambda_i)^2}{(\Sigma\lambda_i)^2 + (\Sigma(1-\lambda_i^2))} = \frac{(\Sigma\lambda_i)^2}{(\Sigma\lambda_i)^2 + \Sigma\delta_i} \quad (5)$$

Where λ_i is the standardized loading of the measurement item "i" on the latent variable and δ is its measurement error. Both statistics are reported by LISREL. Equation (5) is worth remembering because LISREL does not automatically generate this statistic.

Should theory be adequately matched by operationalization in Figure 1, scores on items dealing with, say, integral mathematics, algebra, trigonometry, and mathematical problem solving of story quizzes should be consistent. In other words, we expect a reliable measure in which the score on each of the four examinations varies consistently with the scores of the other three examinations that reflect that type of intelligence. Accordingly, a student who receives an above average score in one test should also receive an equivalently above average score in the other three tests. If, on average, that happens consistently with many students then the scale is judged to have good reliability. Typically composite factor reliability should have the same thresholds as Cronbach's α, above .80.

Reliability, however, only captures part of the psychometric qualities required of a latent variable. Reliability does not capture how the non-common variance of the measurement items correlates and therefore cannot fully assess unidimensionality [Gerbing and Anderson, 1988, Segars, 1997]. In fact, reliability assumes *a priori* that the items are already unidimensional [Green et al., 1977]. As illustrated in Figure 1, reliability analyses will not be able to assess if the non-red areas in the top four items and the non-blue areas in the bottom four items share non-common variance. Reliability analysis thus will miss the shared variance represented by the blue section in items 3 and 4, and the shared variance represented in hues of green. Going back to equations (2) and (3), it becomes obvious why equations (4) and (5), which do not measure the correlations between the non-common variance, cannot assess unidimensionality. Unidimensionality, as defined above, is shown when each measurement item reflects only its associate latent construct without significantly reflecting any other construct. Reliability does not examine how a measurement item of one construct is correlated to a measurement item of another construct. Consequently, Cronbach's α is not a substitute for measuring unidimensionality [Green et al., 1977, Hatti, 1985, Rubio et al., 2001].

CONVERGENT AND DISCRIMINANT VALIDITY

Construct Validity measures the psychometric accuracy of the latent variable by examining its association with other latent variables. Two of the most widely examined aspects of construct validity are *convergent validity* and *discriminant validity*.

Convergent validity examines the magnitude of correlation between item measures of a construct across multiple methods of measurement.

Discriminant validity is the degree of uniqueness achieved from item measures in defining a latent construct [Churchill, 1979].

Together, these properties of validity imply that measures of a latent variable will show a pattern of high correlations within measurement items and a pattern of lower correlations with measurement items that reflect other latent variables [Hair et al., 1998]. Ideally, these characteristics are also realized across measurement methods. This replication across methods is the essence of the MTMM, *Multi-Trait Multi-Method* analysis [Campbell and Fiske, 1959, Churchill, 1979].

In many MIS studies, convergent and discriminant validity are measured through *factorial validity,* i.e., in an exploratory principal components factor analysis (PCA). Convergent and discriminant validity is implied when all the measurement items of each latent variable load with a large coefficient together on the same factor while loading with small coefficients on the factors created by measurement items that reflect the other latent variables. There are distinct differences, however, between a PCA and a CFA. Primary among these differences is that a PCA is exploratory, which means that the measurement items are not assigned *a priori* based on theory and content to a latent construct but rather are assigned statistically based on empirical correlation patterns. Moreover, a PCA, in contrast with a CFA, allows each measurement item to correlate with all the factors [Hair et al., 1998].

In Figure 1, convergent and discriminant validity could be examined in a PCA to verify that the eight measurement items show two eigenvalues above 1, and that the eight items load, when rotated, in such a way that the items dealing with quantitative intelligence load highly on one factor while the items dealing with verbal intelligence load highly on the other. This result would imply convergent validity. In addition, if the items load only with a small coefficient on the other factor, then discriminant validity is also implied.

To some extent this analysis can detect some severe problems in unidimensionality such as the emergence of an unexpected factor or a measurement item that loads highly on a different factor.

However, convergent and discriminant validity as examined through factorial analysis do not specifically examine unidimensionality because the factors are created as a weighted sum of all the measurement items [Gerbing and Anderson, 1988, Hatti, 1985]. Factorial validity as measured through a PCA also has the disadvantage of being unable to examine higher order models, such as a second-order factor analysis [Rubio et al., 2001]. Moreover, exploratory factor analysis considers only a proportion of the variance of each measurement item (shown through the commonality statistics). It ignores the rest of the variance, even though this residual variance may be significantly correlated with that of another measurement item and thus result in a lack of unidimensionality. In addition, by examining the variance shared between each measurement item and the latent variables, an exploratory factor analysis does not examine directly whether the variance of one measurement item is correlated with the variance of another measurement item. (For a further discussion on why PCA is a problematic measure of Unidimensionality, see Hatti et al. [1985] and Schumacker and Beyerlein [2000].)

As a result, using first generation regression models (OLS regression, ANOVA) can result in erroneous conclusions about the relationships among the latent variables (structural model). In these instances, results that are counter to prevailing theory or otherwise equivocal may be attributable to confounds in measurement. This possibility is discussed by Segars and Grover [1993] within the context of Davis' TAM model. In that study, the authors demonstrate underlying problems in measurement that are detected through confirmatory factor analysis (CFA) yet go undetected through traditional analysis. Nonetheless, unless subsequent cross-validation is conducted, this detection may equally be due to capitalization on the randomness of the data providing erroneous conclusions of multidimensionality [Chin and Todd, 1995, MacCallum et al., 1992].

In the following section, we discuss the utility of CFA for assessing measurement and then provide an example of its in Sections 4 and5.

III. THE PRACTICE

While not the only technique for assessing unidimensionality, CFA overcomes many of the limitations inherent in traditional analyses. As noted in Section II, traditional metrics of measurement do not examine the correlation between the non-common variance of the measurement items. LISREL CFA, if modeled as such, can be used to examine all the different variances explicitly, whether it is the common variance specified in a path or other non-common variance [Bollen, 1989, Jöreskog and Sörbom, 1983, Jöreskog and Sörbom, 1989]. Therefore, if amount of shared variance between two measurement items that is not accounted for in the model is significant, the fit indices, especially RMR, and its related Standardized RMR, and the χ^2, will reflect it.

As illustrated in Figure 1, there is a large proportion of shared non-common variance in the IQ example (represented as the non-red areas of items 1 through 4 and the non-blue areas of items 5 through 8). This variance may be ignored by PCA, being an exploratory factor analysis, unless the eigenvalue of some of the non-common variance that is shared across measurement items is large enough to register in the factor analysis. Reliability analysis is also not adequate because the portion of this variance that is shared across latent variables, such as between items 4 and 7 in Figure 1, will be ignored. Reliability regards non-common variance as "noise" and does not examine possible significant correlations between the non-common variance of any of pair of measurement items.

In contrast, the non-common variance is calculated explicitly by covariance-based SEM as residual variance. Any significant correlations among any pair of measurement item residual variance is then examined. When there is a significant degree of shared variance, the fit indices will be reduced significantly unless the shared residual variance is modeled into the loading pattern. This examination of correlation among pairs of residual variance is done regardless of whether the SEM is run as a CFA, as a regression model, or as a path analysis [Bollen, 1989, Jöreskog and Sörbom, 1983, Jöreskog and Sörbom, 1989].

Examining this shared residual variance is at the core of verifying unidimensionality. When the standardized shared residual variance is above 2.58, corresponding to the critical p<.01 threshold, one or both of the measurement items *may* not be unidimensional [Gerbing and Anderson, 1988, Segars, 1997]. Covariance-based SEM automatically reports this analysis of shared residual variance highlighting pairs of measurement items with shared residual variance above 2.58. In addition to this detailed analysis of shared residual variance among pairs of measurement items, the covariance-based SEM provides aggregate measures of possible threats to unidimensionality through the RMR and χ^2 statistics. These two statistics as well as the other statistics that are derived from them directly reflect such possible threats. Specifically, standardized RMR and χ^2 will be larger when such a threat exists.

Based on the shared residual variance statistics, lack of unidimensionality can be reduced by dropping measurement items with a high degree of standardized shared residual variance. However, caution must be taken in this instance to avoid overfitting the model and/or driving the analysis primarily through data rather than theory [Gerbing and Anderson, 1988]. It is important to note here that not every standardized shared residual variance above 2.58 implies a threat to the unidimensionality of the measurement item. As will be shown in the example below, a high degree of shared residual variance between a pair of measurement items often results in a cascading effect where the shared residual variance among other pairs of measurement items also becomes large enough to be significant.

Another set of measures that highlights possible problems in unidimensionality is the modification indices. These statistics examine the approximate change in the overall model χ^2 if a new path is added. A modification index larger than 5 indicates that the inclusion of an additional path will result in a significant improvement in the overall model χ^2. Note that for a χ^2 with one degree of freedom to be significant its value must be at least 5.02.[11]

Finally, as derived from Anderson and Gerbing's [1988] formulation of unidimensionality, lack of unidimensionality may be present when the modification indexes for theta delta and theta epsilon are large. Theta delta is the error component of an X measurement item, while theta epsilon is the error component of a Y measurement item. If all the X and Y measurement items are unidimensional then all the theta delta and theta epsilon statistics should be uncorrelated. If, however, the modification index for one of these is large enough to reduce the overall model χ^2 (i.e., it is above 5.0), then the correlation between the error components of the two measurement items is significant. Therefore, the shared theta epsilon or theta delta variance is due to an unspecified latent variable rather than to random error, implying a violation of unidimensionality. This scenario will be demonstrated in the example in Section IV.

[11] Actually LISREL can be run in a mode where it will free paths one at a time in both the measurement model and in the structural model until the there are no modification indexes larger than 5.0 left. This mode is somewhat analogous to running a stepwise linear regression where one path after another is released until no additional significant change in the F statistic can be achieved. In LISREL this is done with the AD parameter in the OU line. This approach is highly unadvisable because it results not only in over-fitting the model to the data but also in adding paths regardless of their theoretical meaning.

CAVEAT

Not every significant shared covariance indicates a threat to unidimensionality. The standard threshold that many covariance SEM packages apply automatically – shared standardized residuals of at least 2.58, corresponding to a p-value smaller than .01 – is somewhat arbitrary. For example, if the p-value changed from .009 to .011 it may not be a clear indicator that there is no threat to the unidimensionality of measurement items.

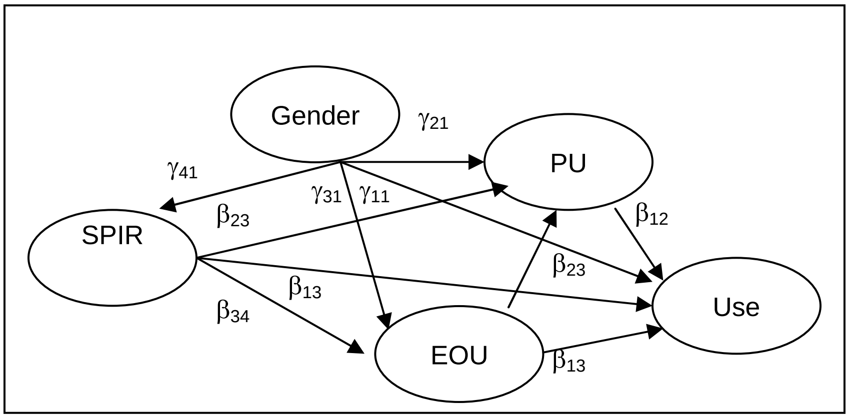
Another problem with determining unidimensionality by examining the size of the standardized residuals and the corresponding modification indexes is that the shared variance may be attributable to random shared non-common variance between the two measurement items. The 2.58 threshold assumes that there is such random shared variance but that the probability of getting such a degree of shared residual between any pair of measurement items by chance alone is less than one in a hundred. Since this randomly occurring shared variance is not related to any unspecified latent variable, assuming a lack of unidimensionality might result in overfitting the model to the sample data. In addition, if a single method is used to collect the data then part of the non-common variance in each measurement item should be shared with other measurement items. This effect is due to possible shared measurement error resulting from the method of questionnaire administration. In many cases, determining the seriousness of violations in accepted thresholds is a matter of judgment that is based on theory. However, when the modification indexes for theta epsilon or theta delta are extremely large (as shown in Model 3 in Section IV), they should not be ignored.

IV. AN ILLUSTRATIVE EXAMPLE

BACKGROUND

The following example illustrates how SEM can be used to examine properties of unidimensionality. The example replicates the model first introduced by Gefen and Straub [1997] in the context of e-mail adoption across cultures and between genders. In general, the model, based on socio-linguistics [Coates, 1986, Tannen, 1994], suggests that women convey and perceive more rapport and compassion in traditional discourse, and should therefore also perceive increased levels of social presence in e-mail. Because social presence is a preferable attribute of communication, it is further hypothesized that it will increase the perceived usefulness of the medium. To examine this model, Gefen and Straub [1997] examined a sample of e-mail users in three airlines.

The example below replicates part of that model but in the context of purchasing flight tickets at an established website. The example uses the original six items of perceived usefulness and the six items of perceived ease of use [Davis, 1989]. Technology acceptance, use, is measured through a scale of purchase intentions at a website [Gefen and Straub, 2000]. Social presence is measured with the original scale used by Gefen and Straub [1997]. The model is shown below in Figure 3. The data were collected from 170 MBA students in a large, urban research university in the North Atlantic region of the US. The students were asked to complete, voluntarily, a questionnaire dealing with Travelocity.com. Complete questionnaires were returned from 160 students. More details about the data are available in Gefen et al. [2000] in CAIS. The questionnaire items are shown in Appendix 2. The mean (ME), standard deviation (SD), Label (LBL), and correlation (KM) files can be derived from Appendix 1. The items in the appendix are labeled EOU for perceived ease of use, PU for perceived usefulness, SPIR for social presence, USE for the intention to use the website, Gender to identify the gender of the respondent (0 = Women, 1 = Men). Gender is a directly observable measure; it is not a latent variable.



Source: Gefen and Straub [1997]

Figure 3: Demonstration Model

MODEL 1

Initially, the model illustrated in Figure 3 was estimated. Appendix 3 contains the LS8 and SE files with the LISREL code. It is worth noting that since no measurement error was expected in the Gender measure, its theta delta was set to zero. The LISREL analysis shows that Gender affects only SPIR significantly; that SPIR significantly affects only Use; that PEOU affects PU; and that PU affects Use. The other paths were insignificant. In all, the SMC (equivalent of an R^2) was 41% for Use and 50% for PU. These results make sense in the case of a website selling flight tickets because EOU is not an integral part of the primary service provided and therefore should not affect USE [Gefen and Straub, 2000]. In addition, there is no interaction with another person in the medium that may make social presence a desired attribute that in turn increases perceived usefulness or ease of use.

When examining the same model in linear regression, equivalent results emerge. Only PU and SPIR affect Use (R^2 = .25, F= 7.63), EOU alone affects PU (R^2 = .32, F= 15.74), and Gender affects SPIR (R^2 = .04, F= 3.91). All the other paths were insignificant. The higher degrees of explained variance in LISREL as shown in the SMC statistic compared with the R^2 values in linear regression are to be expected given that linear regression examines the relationships among scales (latent variables) that are the non-weighted average of their measurement items.

The overall model fit indexes in LISREL are marginally below acceptable thresholds [Gefen et al., 2000]: χ^2_{180} = 314.00 (p-value=0.00), Standardized RMR is 0.053, GFI is .85, NFI is .87, and CFI is .94. Excerpts of the LISREL analysis that are important for the unidimensionality analysis are shown below. The residual analysis suggests some possible threats to unidimensionality, especially in the pairs of items with standardized residuals far above 2.58. These cases can be seen below between EOU5 and EOU6, between PU1 and PU6, and between PU6 and PU5. The pairs are presented in bold font for emphasis below.[12]

---

[12] The bold typeface shown in Figure 4 and subsequent figures is not added by LISREL, it is added here for emphasis.

LARGEST NEGATIVE STANDARDIZED RESIDUALS

RESIDUAL FOR PU3 AND USE1 -2.86

RESIDUAL FOR PU5 AND PU2 -3.67

RESIDUAL FOR PU6 AND PU1 **-3.91**

LARGEST POSITIVE STANDARDIZED RESIDUALS

RESIDUAL FOR USE2 AND USE1 2.66

RESIDUAL FOR PU3 AND PU2 3.25

RESIDUAL FOR PU6 AND PU5 **3.76**

RESIDUAL FOR EOU2 AND PU1 3.04

RESIDUAL FOR EOU5 AND PU1 2.60

RESIDUAL FOR EOU6 AND EOU5 **4.31**

RESIDUAL FOR SPIR2 AND USE3 2.76

RESIDUAL FOR SPIR5 AND USE3 2.95

RESIDUAL FOR GENDER AND USE3 3.44

Figure 4. LIREL Residuals

Threats to unidimensional measurement are also pronounced in the modification indexes (partly shown in Figure 5) Bold typeface added manually to emphasize the values. As noted before, indexes above 5.0 are suspect. Some of these high indexes reflect the same pairs of items shown above, such as between PU1 and PU6, between PU5 and PU6, and between EOU5 and EOU6. Other high indexes show additional threats to unidimensionality, such as between SPIR3 and PU3.

As illustrated, there are some potentially problematic pairs of measurement items. These potential threats were handled by iteratively excluding the most problematic measurement item, one at a time, and then rerunning the analysis to identify the next most problematic measurement item, until there are no more problematic measurement items. The measurement items were discarded one at a time because a high degree of shared residual variance in one pair of measurement items tends to cascade to other pairs. In this manner items USE3, PU6, PU3, and EOU5 were discarded. It is not an uncommon practice to drop items from the original PU and EOU scales, most TAM studies have done so [Gefen and Straub, 2000]. It is important to realize when going through the process that:

As the sample size increases so does the p-value of the shared residuals.

As a result, a pair of measurement items that does not have significant shared residual variance with a relatively small sample size may have significant shared residual variance with a larger sample size.[13]

MODIFICATION INDICES FOR THETA-EPS

	USE1	USE2	USE3	PU1	PU2	PU3
	--------	--------	--------	--------	--------	--------
USE2	**7.08**	- -				
PU3	**8.98**	1.53	0.58	0.66	**10.56**	- -
PU5	3.88	0.77	0.02	0.03	**13.48**	0.34
PU6	2.27	0.01	5.84	**15.31**	0.03	0.03
SPIR2	0.00	0.03	0.03	1.13	2.12	**7.21**
SPIR3	0.35	0.59	0.30	3.29	3.52	**13.18**

	PU4	PU5	PU6	EOU1	EOU2	EOU3
	--------	--------	--------	--------	--------	--------
PU6	0.01	**14.10**	- -			
SPIR5	**10.10**	0.01	1.56	0.93	0.21	4.56

	EOU4	EOU5	EOU6	SPIR1	SPIR2	SPIR3
	--------	--------	--------	--------	--------	--------
EOU6	0.32	**18.60**	- -			

Figure 5. Modification Indices

MODEL 2

The resulting model after dropping these four items showed no overt threats to unidimensionality. Appendix 4 contains the LS8 and SE files with the LISREL code. The model shows the same pattern of significant paths as in Model 1, albeit with much better fit indexes: χ^2_{110} = 134.46 (p-value=0.057), Standardized RMR is 0.039, GFI is .92, NFI is .92, and CFI is .98. The SMC was 10% lower for Use at 37% and 10% higher for PU at 55%. There are no standardized residuals greater than 2.58 and just one modification index slightly above 5.0. These results suggests that the high degrees of shared residual variance in the measurement items that were dropped in Model 1 caused additional pairs of measurement items to show high degrees of residual

[13] The same applies to all the p-values in the model, including those of the paths between the latent variables. In the model analyzed here, if the sample size were doubled, then the paths from EOU to USE and from SPIR to PU and to EOU would become significant, totally changing the conclusions that might be drawn from the analysis.

variance, as in Figure 4. The equivalence pattern of significant paths in Model 1 and in Model 2 also suggests that the possible threats to unidimensionality were apparently not serious enough to actually change the model. Replicating the analysis in linear regression with the revised scales produced equivalent results: still only PU and SPIR affect Use (R^2 = .25, F= 11.11), and only EOU affects PU (R^2 = .46, F= 39.60).

MODEL 3

To demonstrate a serious threat to unidimensionality and its impact on a pattern of significant paths, a new variable SPIR23 was included in the analysis. SPIR23 was created as the square of the mean of SPIR2 and SPIR3 in the data points and was then multiplied by 2.6 to make its range the same 1 to 7 as the other measurement items. Created in this manner, SPIR23 should exhibit a lack of unidimensionality with SPIR2 and with SPIR3. This model was run without items USE3, PU3, PU6, and EOU5. Appendix 5 contains the LS8 and SE files with the LISREL code.

The analysis shows that the pattern of significance paths changed markedly. Now, SPIR suddenly affects PU while Gender no longer affects SPIR. In other words, the conclusions of the research might have been very different had this item been included. Also, the fit indexes are now worse, reflecting the unidimensionality problem created by this new item. However, the fit indices are still close enough to accepted thresholds that without explicitly looking for threats to unidimensionality a mistaken conclusion could have been reached. The fit indexes are marginal: χ^2_{126} = 352.08 (p-value=0.0), Standardized RMR is 0.057, GFI is .83, NFI is .85, and CFI is .90. However, looking for threats to unidimensionality in the standardized residuals shows many pairs greater than 2.58 (these pairs are marked in Figure 6 in bold typeface for emphasis), meaning that the model is unstable and may lack unidimensionality in these measurement items:

LARGEST NEGATIVE STANDARDIZED RESIDUALS

RESIDUAL FOR SPIR3 AND SPIR2 -11.61

RESIDUAL FOR SPIR23 AND SPIR2 -8.63

RESIDUAL FOR SPIR23 AND SPIR3 -9.00

LARGEST POSITIVE STANDARDIZED RESIDUALS

RESIDUAL FOR SPIR2 AND SPIR1 4.19

RESIDUAL FOR SPIR3 AND SPIR1 4.55

RESIDUAL FOR SPIR4 AND SPIR1 2.98

RESIDUAL FOR SPIR5 AND SPIR1 5.91

RESIDUAL FOR SPIR5 AND SPIR2 3.73

RESIDUAL FOR SPIR5 AND SPIR3 3.06

RESIDUAL FOR SPIR5 AND SPIR4 4.04

RESIDUAL FOR SPIR23 AND SPIR1 7.10

RESIDUAL FOR SPIR23 AND SPIR4 3.45

RESIDUAL FOR SPIR23 AND SPIR5 6.13

RESIDUAL FOR GENDER AND SPIR3 2.69

Figure 6. Largest Negative Standardized Residuals; SPIR23 Included in Model

The pattern in Figure 6 illustrates a key point. The largest group of shared residuals is among SPIR2, SPIR3, and SPIR23, which would be expected given that SPIR23 was deliberately created to have shared variance with SPIR2 and SPIR3. It is also worth noting that these high degrees of shared residual variance caused a cascading effect in which other pairs of items have higher degrees of shared residual variance.

MODIFICATION INDICES FOR THETA-EPS

	USE2	USE1	PU1	PU2	PU4	PU5
	----	----	----	----	----	----
USE2	- -					
USE1	- -	- -				
PU1	1.72	1.77	- -			
PU2	0.00	0.10	0.07	- -		
PU4	0.51	3.91	0.40	2.25	- -	
PU5	0.63	3.64	1.54	1.77	1.68	- -
EOU1	0.15	0.25	0.16	0.34	1.21	0.52
EOU2	1.74	0.10	2.16	0.37	0.07	0.18
EOU3	0.74	1.20	1.76	2.25	3.50	0.52
EOU4	0.05	0.09	3.23	5.58	1.65	1.31
EOU6	0.89	0.06	0.00	0.94	0.71	0.24
SPIR1	0.26	1.63	0.34	1.52	2.44	0.16
SPIR2	0.06	2.96	5.41	2.00	2.18	0.09
SPIR3	0.27	1.75	0.00	6.04	4.42	0.58
SPIR4	0.19	0.57	2.83	0.26	4.35	0.01
SPIR5	0.01	4.83	3.02	0.10	0.59	0.48
SPIR23	0.00	4.61	2.78	0.36	4.12	0.10

	EOU1	EOU2	EOU3	EOU4	EOU6	SPIR1
	----	----	----	----	----	-----
EOU1	- -					
EOU2	0.76	- -				
EOU3	1.44	1.15	- -			
EOU4	4.72	2.55	0.09	- -		
EOU6	0.30	0.40	1.18	5.24	- -	
SPIR1	0.24	0.10	0.10	0.07	5.85	- -
SPIR2	0.00	0.05	0.03	0.16	2.78	17.52
SPIR3	0.12	0.24	0.07	0.32	3.58	20.71
SPIR4	1.78	1.10	0.14	0.71	0.26	8.89
SPIR5	0.68	0.37	3.02	0.23	5.98	34.92
SPIR23	0.00	0.64	0.35	0.09	6.02	**50.47**

	SPIR2	SPIR3	SPIR4	SPIR5	SPIR23
	-----	-----	-----	-----	------
SPIR2	- -				
SPIR3	134.71	- -			
SPIR4	5.02	2.34	- -		
SPIR5	13.93	9.36	16.30	- -	
SPIR23	**74.47**	**81.07**	11.92	**37.58**	- -

Figure 7. Modification Indices for Theta-Epsilon

Even more pronounced is the impact on the modification indices. Here, the higher statistics associated with SPIR23 suggest that shared variance in this item is not due to random error (unlike other instances of shared variance). This finding is hardly surprising given that the item was created for the purpose of demonstrating this effect. Figure 7 shows the modification indices for theta epsilon. The bold typeface was added here to emphasizing these high indexes.

V. CONCLUSION

The unidimensionality of measurement items is a crucial attribute of any latent variable (scale). In traditional statistical tools, unidimensionality is assumed to exist [Anderson and Gerbing, 1988, Segars, 1997]. When the threat of unidimensionality is minor, as in Model 1, ignoring this threat does not markedly alter the structural model and does not result in changes in the pattern of significant paths among the latent variables. However, as illustrated in Model 3, a serious violation of unidimensionality may result in false or equivocal conclusions. In this scenario, researchers may begin to question or re-examine theoretical conventions resulting in a fragmented and inconclusive line of inquiry. This danger is particularly acute as the issues of interest become more complex and require more robust representations of theoretical concepts.

Based on work by Gerbing and Anderson [1988], the present tutorial developed a conceptual and operational definition of unidimensionality and demonstrated potential pitfalls in modeling items that do no exhibit this statistical property. Utilizing SEM, we also demonstrate methods for detecting violations of unidimensionality and correcting these instances. Importantly, SEM is neither the only nor necessarily the best method for analyzing unidimensionality. SEM involves its own set of statistical assumptions that confine its applicability. Further, alternative approaches such as Item Response Theory may be more applicable and more effective in certain research contexts [Hambleton et al., 1991]. Clearly, future research efforts should assess alternative approaches along with their strengths/weaknesses relative to SEM. The important point of this tutorial is to surface the issue and offer a means of addressing its impact, in the hope of building further credibility and consistency within the domain of IS research.

Editor's Note: This article was received on September 26, 2001. It was with the author approximately 3 months for 3 revisions. It was published on July 10, 2003.

ACKNOWLEDGEMENT

Special thanks to Al Segars of the Kenan-Flagler Business School at the University of North Carolina at Chapel Hill for assistance and helpful review of an early version of this tutorial.

Editor's Note: This article was received on September 26, 2001. It was with the author approximately three months for three revisions. It was published on July10, 2003.

REFERENCES

Anderson, J. C. and D. W. Gerbing (1988) "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin* (103) 3 (Fall), pp. 411-423.

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

Boudreau, M., D. Gefen, and D. W. Straub (2001) "Validation in IS Research: A State-of-the-Art Assessment," *MIS Quarterly* (25) 1, pp. 1-16.

Campbell, D. T. and D. W. Fiske (1959) "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin* (56) 2 (March), pp. 81-105.

Chin, W. W. and P. A. Todd (1995) "On the Use, Usefulness, and Ease of Use of Structural Equation Modeling in MIS Research: A Note of Caution," *MIS Quarterly* (19) 2, pp. 237-246.

Churchill, G. A., Jr. (1979) "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research* (16) 1 (February), pp. 64-73.

Coates, J. (1986) *Women, Men and Languages: Studies in Language and Linguistics*. London, UK: Longman.

Cronbach, L. J. (1951) "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika* (16) September, pp. 297-334.

Davis, F. D. (1989) "Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology," *MIS Quarterly* (13) 3 (September), pp. 319-340.

Gefen, D. and D. Straub (2000) "The Relative Importance of Perceived Ease-of-Use in IS Adoption: A Study of e-Commerce Adoption," *JAIS* (1) 8, pp. 1-30.

Gefen, D. and D. W. Straub (1997) "Gender Differences in Perception and Adoption of E-Mail: An Extension to the Technology Acceptance Model," *MIS Quarterly* (21) 4 (December), pp. 389-400.

Gefen, D., D. W. Straub, and M. Boudreau (2000) "Structural Equation Modeling And Regression: Guidelines For Research Practice," *Communications Of The Association For Information Systems* (4) 7, pp. 1-70.

Gerbing, D. W. and J. C. Anderson (1988) "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment," *Journal of Marketing Research* (25) May, pp. 186-192.

Gorden, R. L. (1977) *Unidimensional Scaling of Social Variabls: Concpts and Procedures*. New York: The Free Press.

Green, V. C., R. W. Lissitz, and S. A. Mulaik (1977) "Limitations of Coefficient Alpha as an Index of Test Unidimensionality," *Educational and Psychometric Measurement* (37), pp. 827-838.

Hair, J. F., Jr., R. E. Anderson, R. L. Tatham, and W. C. Black (1998) *Multivariate Data Analysis with Readings, 5th Edition*. Englewood Cliffs, NJ: Prentice Hall.

Hambleton, R., H. Swaminathan, and H. Swaminathan (1984) *Item Response Theory: Principles and Applications*: Kluwer Academic Publishers.

Hambleton, R. K., H. Swaminathan, and H. J. Rogers (1991) *Fundamentals of Item Response Theory*. New York, NY: Sage Publications.

Hatti, J. (1985) "Methodology Review: Assessing Unidimensionality of Tests and Items," *Applied Psychological Meeasurement* (9) 2, pp. 139-164.

Hatti, J. (1997) "Common Problems in Structural Modeling," University of Auckland, http://www.arts.auckland.ac.nz/edu/staff/jhattie/struct_intro.html.

Hatti, J., K. Krakowski, H. J. Rogers, and H. Swaminathan (1996) "An Assessment of Stout's Index of Essential Unidimensionality," *Applied Psychological Measurement* (20) 1, pp. 1-14.

Hox, J. J. and T. M. Bechger (1998) "An Introduction to Structural Equation Modeling," *Family Science Review* (11), pp. 354-373.

Jöreskog, K. G. and D. Sörbom (1983) *LISREL: Analysis of Linear Structural Relations by the Method of Maximum Likelihood*, 2nd Edition edition. Chicago: National Educational Resources.

Jöreskog, K. G. and D. Sörbom (1989) *LISREL7: A Guide to the Program and Applications*, 2nd edition. Chicago: SPSS Inc.

Jöreskog, K. G. and D. Sörbom (1993) *LISREL8: Structural Equation Modeling with SIMPLIS Command Language*, 2nd edition. Chicago, IL: Scientific Software International.

Lord, F. M. and M. R. Novick (1968) *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

MacCallum, R. C., M. Roznowski, and L. B. Necowitz (1992) "Model Modifications in Covariance Structure Analysis: The Problem of Capitalization on Chance," *Psychological Bulletin* (111), pp. 490-504.

McDonald, R. P. (1979) "The Structural Analysis of Multivariat Data: A Sketch of a General Theory," *Multivariate Bhavioral Research* (14), pp. 21-28.

McDonald, R. P. (1981) "The Dimensionality of Tests and Items," *British Journal of Mathematical and Statistical Psychology* (34), pp. 100-117.

McIver, J. P. and E. G. Carmines (1982) *Unidimensional Scaling*. Vol. 24. Beverly Hills: Sage.

Nunnally, J. C. and I. H. Bernstein (1994) *Psychometric Theory, Third Edition*. New York: McGraw-Hill.

Peterson, R. A. (1994) "A Meta-analysis of Cronbach's Coefficient Alpha," *Journal of Consumer Research* (21) September, pp. 381-391.

Rubio, D. M., M. Berg-Weger, and S. S. Tebb (2001) "Using Structural Equation Modeling to Test for Multidimensionality," *Structural Equation Modeling* (8) 4, pp. 613-626.

Schumacker, R. E. and S. T. Beyerlein (2000) "Confirmatory Factor Analysis With Different Correlation Types and Estimation Methods," *Structured Equation Modeling* (7) 4, pp. 629-636.

Segars, A. (1994) “Scale Development in Information Systems Research: A Paradigm Incorporating Unidimensionality and Its Assessment.” *International Conference on Information Systems (ICIS), Cleveland, Ohio, 1994*.

Segars, A. H. (1997) "Assessing the Unidimensionality of Measurement: A Paradigm and Illustration within the Context of Information Systems Research," *Omega* (25) 1 (February), pp. 107-121.

Segars, A. H. and V. Grover (1993) "Re-Examining Perceived Ease of Use and Usefulness: A Confirmatory Factor Analysis," *MIS Quarterly* (17) 4 (December), pp. 517-525.

Stout, W. F. (1987) "A Nonparametric Approach for Assessing Latent Trait Unidimensionality," *Psychometrika* (52), pp. 589-617.

Tannen, D. (1994) *You Just Don't Understand Women and Men in Conversation*. New York, NY: Ballantine Books.

Werts, C. E., R. L. Linn, and K. G. Jöreskog (1974) "Interclass Reliability Estimates: Testing Structural Assumptions," *Educational and Psychological Measurement* (34), pp. 25-33.

BIBLIOGRAPHY

Additional tutorials and references of interest are cited here. References are included in the list of references. A good discussion of SEM, including a discussion on CFA in comparison with PCA is given by Hox and Bechger [1998]. Another good discussion about the differences between an exploratory factor analysis and CFA is given by Rubio et al. [2001], including an example with empirical data. An excellent discussion about confirmatory factor analysis and its application to different data types is given by Schumacker and Beyerlein [2000] and by Hair et al. [1998]. A good discussion of the differences between unidimensional and multi-dimensional scaling is given by McIver and Carmines [1982]. A discussion of the evolution of scale theory and how unidimensionality was assessed prior to the establishment of statistical tests in the context of early scales is given by Gorden [1977]. Another excellent source on SEM is Hatti's webpage on "Common Problems in Structural Modeling" [Hatti, 1997].

APPENDIX I. THE DATA (THE KM, SD, ME, LBL FILES)

Label	Mean	Std.	Correlation Table										
EOU1	2.16	1.05	1.0000										
EOU2	2.21	1.20	.6353	1.0000									
EOU3	2.06	1.07	.6944	.7284	1.0000								
EOU4	2.43	1.20	.5768	.6313	.7342	1.0000							
EOU5	2.32	1.17	.5995	.6451	.7067	.7028	1.0000						
EOU6	2.20	1.06	.6377	.6751	.7417	.7425	.8155	1.0000					
Gender	.61	.49	.0669	.1656	.0552	.1644	.1181	.0315	1.0000				
PU1	2.31	1.40	.4841	.5416	.5382	.4731	.5421	.5161	.2055	1.0000			
PU2	2.67	1.41	.4353	.4301	.4222	.5159	.4642	.4321	.1339	.5898	1.0000		
PU3	2.56	1.25	.3335	.3775	.3935	.4052	.3947	.3632	.0007	.4958	.6719	1.0000	
PU4	2.48	1.25	.4413	.4916	.4752	.5327	.5136	.5190	.1609	.6435	.6714	.6457	1.0000
PU5	2.49	1.30	.4888	.4919	.5212	.4894	.5443	.5209	.1335	.5948	.5825	.6106	.7056
PU6	2.53	1.28	.4820	.4530	.4538	.5174	.5105	.4777	.1193	.5302	.6832	.6397	.7487
SPIR1	3.44	1.75	.1656	.0843	.1076	.1326	.0201	.0189	.2151	.1386	.0953	.0584	.2073
SPIR2	3.07	1.70	.2187	.1232	.1594	.1978	.1224	.1106	.0847	.1318	.1951	.1724	.2138
SPIR3	3.79	1.67	.2161	.1161	.1463	.1642	.1150	.0922	.2262	.1868	.1031	.0036	.2265
SPIR4	4.45	2.77	.0154	.0759	.0364	.0299	.0293	.0592	.1354	.0196	.0594	-.0155	.1647
SPIR5	4.20	1.68	.2025	.1136	.1646	.1348	.0590	.0449	.1455	.0632	.1154	.0736	.1019
USE1	4.23	1.63	.2951	.2946	.3131	.2748	.2247	.2834	.0547	.3099	.3560	.1848	.3249
USE2	4.88	1.72	.2238	.2749	.2245	.2214	.2577	.2058	.0016	.3652	.3469	.3012	.3664
USE3	4.67	1.78	.2991	.2838	.2670	.2829	.3382	.3133	.2452	.3399	.3788	.2697	.3073
SPIR23	4.67	1.63	.2450	.1419	.1801	.2080	.1504	.1331	.1649	.1795	.1645	.1054	.2215

Label	Correlation Table										
PU5	1.0000										
PU6	.7900	1.000									
SPIR1	.1935	.1514	1.0000								
SPIR2	.2061	.2283	.7201	1.0000							
SPIR3	.2297	.2110	.7294	.7263	1.0000						
SPIR4	.1013	.1096	.4224	.4296	.4043	1.0000					
SPIR5	.1641	.1733	.7297	.7487	.7317	.4898	1.0000				
USE1	.4303	.3942	.2674	.3004	.2578	.1730	.2863	1.0000			
USE2	.3751	.3733	.2259	.2462	.1967	.1507	.2009	.6452	1.0000		
USE3	.3624	.2919	.3927	.4127	.4002	.3205	.4236	.5576	.5262	1.0000	
SPIR23	.2333	.2273	.7609	.9244	.9198	.4503	.7947	.2872	.2363	.4300	1.000

APPENDIX II. THE QUESTIONNAIRE ITEMS

Questionnaire items were measured on a 7 point scale ranging from strongly agree through neutral to strongly disagree.

Code	
EOU1	Travelocity.com is easy to use
EOU2	It is easy to become skillful at using Travelocity.com
EOU3	Learning to operate Travelocity.com is easy
EOU4	Travelocity.com is flexible to interact with
EOU5	My interaction with Travelocity.com is clear and understandable
EOU6	It is easy to interact with Travelocity.com
PU1	Travelocity.com is useful for searching and buying flights
PU2	Travelocity.com improves my performance in flight searching and buying
PU3	Travelocity.com enables me to search and buy flights faster
PU4	Travelocity.com enhances my effectiveness in flight searching and buying
PU5	Travelocity.com makes it easier to search for and purchase flights
PU6	Travelocity.com increases my productivity in searching and purchasing flights
SP1	There is a sense of human contact in the Web-site
SP2	There is a sense of personalness in the Web-site
SP3	There is a sense of sociability in the Web-site
SP4	There is a sense of human warmth in the Web-site
SP5	There is a sense of human sensitivity in the Web-site
USE1	I am very likely to buy books from Travelocity.com
USE2	I would use my credit card to purchase from Travelocity.com
USE3	I would not hesitate to provide information about my habits to Travelocity

APPENDIX III:LS8 MODEL 1

```
DA NI=22 NO=160
LA fi=demo.lbl
KM FI=demo.km
ME fi=demo.me
SD fi=demo.sd
SE fi=uni.se
MO NX=1 NK=1 NY=20 NE=4 BE=FU,FI GA=FU,FI
fr ly 2 1 ly 3 1  /* Use
fr ly 5 2 ly 6 2 ly 7 2 ly 8 2 ly 9 2    /* PU
fr ly 11 3 ly 12 3 ly 13 3 ly 14 3 ly 15 3  /* EOU
fr ly 17 4 ly 18 4 ly 19 4 ly 20 4  /*SPIR
va 1 ly 1 1 ly 4 2 ly 10 3  ly 16 4

fi td 1 1  /* Gender
va 1 lx 1 1
va 0 td 1 1

fr be 1 2 be 2 3 be 1 3    /* TAM: EOU -> PU -> Use
fr be 1 4 be 2 4 be 3 4    /* SPIR effects
fr ga 1 1 ga 2 1 ga 3 1 ga 4 1 /* Gender effects

OU MI RS EF MR SS SC
```

SE Model 1

```
USE1
USE2
USE3
PU1
PU2
PU3
PU4
PU5
PU6
EOU1
EOU2
EOU3
EOU4
EOU5
EOU6
SPIR1
SPIR2
SPIR3
SPIR4
SPIR5
GENDER
/
SPIR23
```

APPENDIX IV. LS8 MODEL 2

```
DA NI=22 NO=160
LA fi=demo.lbl
KM FI=demo.km
ME fi=demo.me
SD fi=demo.sd
SE fi=uni1.se   /* the SE file sorts and discards items
MO NX=1 NK=1 NY=16 NE=4 BE=FU,FI GA=FU,FI
fr ly 2 1 /*  Use
fr ly 4 2 ly 5 2 ly 6 2     /* PU
fr ly 8 3 ly 9 3 ly 10 3 ly 11 3  /* EOU
fr ly 13 4 ly 14 4 ly 15 4 ly 16 4 /*SPIR
va 1 ly 1 1 ly 3 2 ly 7 3  ly 12 4

fi td 1 1  /* Gender
va 1 lx 1 1
va 0 td 1 1

fr be 1 2 be 2 3 be 1 3    /* TAM: EOU -> PU -> Use
fr be 1 4 be 2 4 be 3 4 /* SPIR effects
fr ga 1 1 ga 2 1 ga 3 1 ga 4 1 /* Gender effects

OU MI RS EF MR SS SC
```

SE Model 2

```
USE2
USE1

PU1
PU2
PU4
PU5

EOU1
EOU2
EOU3
EOU4
EOU6

SPIR1
SPIR2
SPIR3
SPIR4
SPIR5

GENDER
/
EOU5
PU6
PU3
USE3
SPIR23
```

Appendix V. LS8 Model 3

```
DA NI=22 NO=160
LA fi=demo.lbl
KM FI=demo.km
ME fi=demo.me
SD fi=demo.sd
SE fi=uni2.se
MO NX=1 NK=1 NY=17 NE=4 BE=FU,FI GA=FU,FI
fr ly 2 1 /*  Use
fr ly 4 2 ly 5 2 ly 6 2     /* PU
fr ly 8 3 ly 9 3 ly 10 3 ly 11 3  /* EOU
fr ly 13 4 ly 14 4 ly 15 4 ly 16 4 /*SPIR
fr ly 17 4  /* SPIR23 where there should be unidimensionality threats
va 1 ly 1 1 ly 3 2 ly 7 3  ly 12 4

fi td 1 1  /* Gender
va 1 lx 1 1
va 0 td 1 1

fr be 1 2 be 2 3 be 1 3    /* TAM: EOU -> PU -> Use
fr be 1 4 be 2 4 be 3 4 /* SPIR effects
fr ga 1 1 ga 2 1 ga 3 1 ga 4 1 /* Gender effects

OU MI ME=ML RS EF MR SS SC
```

SE Model 2

```
USE2
USE1

PU1
PU2
PU4
PU5

EOU1
EOU2
EOU3
EOU4
EOU6

SPIR1
SPIR2
SPIR3
SPIR4
SPIR5

SPIR23

GENDER
/
EOU5
PU6
PU3
USE3
```

Communications of the Association for Information Systems

ISSN: 1529-3181