[Practical Assessment, Research, and Evaluation](https://scholarworks.umass.edu/pare)

Volume 17 *Volume 17, 2012* Article 8

2012

A Demonstration of a Systematic Item-Reduction Approach Using Structural Equation Modeling

Karen Larwin

Milton Harvey

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 17, Number 8, April 2012 ISSN 1531-7714

A Demonstration of a Systematic Item-Reduction Approach Using Structural Equation Modeling

Karen Larwin
Youngstown State University

Milton Harvey
Kent State University

Establishing model parsimony is an important component of structural equation modeling (SEM). Unfortunately, little attention has been given to developing systematic procedures to accomplish this goal. To this end, the current study introduces an innovative application of the jackknife approach first presented in Rensvold and Cheung (1999). Unlike the traditional application of jackknife procedures for the purpose of identifying outliers and influential cases within a dataset, this jackknife procedure is applied for the purpose of identifying and eliminating items from a structural model. Items are identified through the jackknife procedure and eliminated from the model without altering the measurement or structural integrity of the model. The goal of this application is to create the most parsimonious model by reducing the number of items in an inventory, without altering the construct represented by the model.

The creation of shorter versions of scales in psychology and allied fields is fairly common. For example in developing a shorter version of WISC-III for clinical use, Donder's (1997, p. 15), goal was to develop "A short form that would maintain the desirable psychometric properties of the full WISC-III (in terms of factor structure and reliability and validity of the instrument)." In other instances, a new shorter model was developed because the existing shorter version did not maintain the conceptual and measurement integrity of the original longer form (see Jackson, Furnham, Forde and Cotter, 2000 for an assessment of the shorter version of the Eysenck Personality Profiler). It was for this reason that Petrides, Jackson, Furnham and Levine (2004) used CFA in developing an "improved new version" (p. 222). Another goal for developing shorter versions of scales was because "reduction of items was to achieve a higher acceptability of the questionnaire in the population, aiming for shorter times of administration, better response rates and lower rates of missing data" (Grossi, Groth, Mosconi, Cerutti, Pace, Compare and Apolone (2006, p. 89). Grossi *et al*. (2006) administered the original 22-item health-related Psychological General Well-Being Index, PGWBI (Italian Version) to 1,015 to a "representative sample … of Italy dwelling Italians." Using the summary scores as a dependent variable and the 22 items as independent variables they used stepwise regression and identified six items that accounted for at least 90 percent of the variance in the summary scores. This resulted in the six-item PGWB-S.

A number of studies have reduced items through exploratory factor analytic approaches (e.g., Clark & Goldsmith, 2006; Salzberger, 2006). Specifically, exploratory factor analysis has been used to identify items with loadings below 0.4 on any of the theorized factors. These items are eliminated from the model. This method is problematic in that it does not take into account the structure of the original factors or the structure of the model. Additionally, this approach incorporates exploratory factor analyses which should only be used with the *exploring* of new inventories, so its generalization to existing scales is not a valid application.

Benson & Bandalos (1992) attempted to reduce the number of items in the forty-item Reaction To Tests (RTT) inventory. The RTT (Sarason, 1984) is a forty-item inventory used to assess test anxiety by measuring the participant's tension, worry, test-irrelevant thinking and bodily reactions. Specifically, the RTT was modeled with a second order factor structure in which four subscales (tension, worry, test irrelevant thinking, and bodily symptoms) directly explained the original forty items, while test anxiety was a higher-order construct explaining the four first-order constructs. The goal of their investigation was to demonstrate that a shorter version of the scale could be developed while maintaining the four-factor model, with the "…same degree of precision as the original scale" (p.643). After conducting confirmatory factor analysis in an effort to demonstrate that the forty-items supported a four-factor model, Benson and Bandalos (1992) reported that they deleted items which had duplicate wording with other questions, or items that had large modification indices or large standardized residuals. Unfortunately, they did not demonstrate that the shorter version maintained the structural integrity of the original, larger version. The authors also failed to explain how decisions regarding standardization residuals and modification indices were made.

The goal of this paper is to demonstrate the use of a systematic jackknife approach, in an effort to produce a shorter version of the Statistics Self-Perception scale (see Larwin, 2007). The theorized factor structure, when tested, demonstrated very good fit. However, due to the size and complexity of the model, it seemed that any attempt at creating a shorter version of any scale by reducing the number of items in the model(s) should establish model parsimony while resulting in a model that computationally is equally valid.

Establishing Model Parsimony

Structural equation modeling (SEM) has grown in popularity over the last thirty years, however not much attention has focused on developing a systematic heuristic for establishing model parsimony. The present investigation employs jackknifing to develop an effective approach to item reduction that results in a more parsimonious model that maintains the integrity of the original structural model. Model parsimony in understanding psychological constructs and in computing structural models is important on a number of levels. According to the principle of parsimony, the explanation of any psychological construct or phenomenon should make as few assumptions as possible, eliminating any items or factors that make no difference in the observable predictions or explanation of a theory or hypothesis (Epstein, 1984). Regarding SEM, Bollen (1989) and Hayduk (1987) maintain that if it is judged that more than one model appropriately fits the data, while also supporting the original theory of the structural model, the most parsimonious (the simplest) model should be selected (Bollen, p. 72).

Additionally, increased model complexity can increase the probability of *catastrophic cancellation*. Specifically, catastrophic cancellation is the result of rounding errors in computer programs. This probabilistically increases as very small quantities of numbers are subjected to arithmetic operations which are computed from larger quantities of numbers (Hanson, 2007). The result of catastrophic cancellation is a loss of precision, specifically in the computation of the elements of one of some eleven matrices used in SEM. Therefore, reducing the number of items in the model assists in the precision of computations as well as model estimation.

Catastrophic cancellation is one of the consequences of multicollinearity that occurs when items are so highly correlated that it becomes difficult to distinguish their individual influences; the departure from orthogonality in a set of independent variables. Exact multicollinearity occurs when there is linear dependence within a set of independent variables and the associated matrix of inter-correlations is singular.

In structural equation modeling, multicollinearity may cause computational problems including non-convergence. "Without convergence, one has no leverage to evaluate the goodness of the model or parameter estimates …with existing procedures in SEM. When the sample covariance matrix **S** is literally singular, existing procedures in SEM do not permit calculations of statistics for the overall model evaluation" (Yuan and Chan, 2008, p. 4843). Even if there is convergence, the results may be untenable because of the following possibilities: (1) the existence of some negative variance estimates – the Heywood cases (Heywood, 1931; Rindskopf, 1984; Dillon, Kumar & Mulani, 1987; Chen, Bollen, Paxton, Curran, & Kirby, 2001); (2) improper parameter estimates. "The population parameter may be a value that is acceptable but close to the boundary of admissible values" (Bollen, 1989, p. 282); (3) "a model can fit perfectly yet be associated with problematic lower order components , such as parameter estimates that are biased, small in magnitude, or opposite to theoretical expectations (Tomarken & Waller, 2005, p. 50), and (4) multicollinearity, which tends to cause increases in the standard errors of coefficients of the affected (collinear) variables and the increased standard errors in turn mean that coefficients for some independent variables may be statistically insignificant, leading to inference errors.

Some scholars seem to contend that structural equation models are robust enough that multicollinearity is not an estimation issue (Malhotra, Peterson, & Kleiser, 1999; Verbeke & Bagozzi, 2000). However, as Marsh, Dowson, Pietsch and Walker (2004, p. 518) observed, "the use of sophisticated statistical tools such as structural equation modeling (SEM) can mislead researchers into thinking that such well-known problems are no longer relevant." Many other scholars echo this point of view (Freedman, 1987; Malhotra, Peterson & Kleiser, 1999; Pietsch, Walker, & Chapman, 2003). Even sophisticated methods like structural equation and complex regression models are adversely affected by multicollinearity. "Multicollinearity is a ubiquitous phenomenon that can produce strange, misleading, or uninterpretable results when a set of highly related independent variables is used to predict a dependent variable. At least the detection and consequences - if not the resolution - of multicollinearity problems are well understood in traditional analyses of manifest (non-latent) variables" (Marsh, Dowson, Pietsch, & Walker, 2004, p. 518). Similar issues are raised in allied fields such as economics. For example, Mela and Kopalle (2002, p. 667) maintained that "The problem of collinearity in empirical research is among the most endemic concerns raised by marketers. In fact, a recent search in EconLit revealed 154 studies discussing collinearity or multicollinearity in their abstracts. A similar full text search of *Applied Economics* (using Infotrac) yielded 220 articles since 1991."

The above discussion illustrates the fact that like regression SEM models are equally affected by high multicollinearity, but when is this the case? As Grewal, Cote, and Baumgartner (2004, p. 520) observed, researchers may ignore multicollinearity because of practical considerations. Existing guidelines about when multicollinearity is likely to cause problems are often ambiguous, procedures for mitigating multicollinearity are frequently of limited usefulness and, most importantly, little is known about how to deal with multicollinearity in the context of SEM. The best solution would be to avoid multicollinearity problems in the first place.

As discussed earlier, in many research situations where the original unidimensional factor consists of a very large number of items, using many methods, researchers reduced the number of items in a model without full consideration of the integrity of the original larger construct. "If construct validity is supported by confirmation of a hypothesized dimensional structure, other types of scale refinement or assessment may be considered" (MacCallum & Austin, 2000, p. 208). One such refinement is the goal of this paper.

The present investigation explores an approach to item reduction and parsimony using an application of jackknifing. In their attempt to eliminate records from their data set, Rensvold and Cheung (1999) used a jackknife approach with LISREL to identify influential outliers in their data. They ordered the models according to the CFI indices and deleted the records with the lowest CFI-values; records were removed one case at a time. The procedure developed here is adapted from a study by Rensvold and Cheung (1999). Although they developed the jackknife approach for the purpose of reducing records in a SEM, rather than items, we believe some variation of their approach might be a viable means of achieving model parsimony for variables instead of cases. The current investigation employs a similar jackknife approach. Analogous to the jackknife application of Rensvold and Cheung (1999), the full model in the present study was estimated using all the variables, then estimated again and again with items excluded one at a time. The impact of item deletion on the model's structure was evaluated after successive iterations. Although computationally intensive, the resultant reduced model maintained the integrity of the construct.

For the current investigation, two different models were used to assess the viability of this jackknife approach: (i) a unidimensional congeneric model of Statistics Related Self-Efficacy, with 14 original items and (ii) a Multi-Dimensional Model of Statistics-Related Anxiety. The unidimensional congeneric model, conceptualized as a first order model, is the simplest model in SEM, and was the logical starting point for this methodological investigation. This model is summarized next.

The First-Order Structural Equation Model

The first-order measurement model is concerned with the variance shared by directly measured observed variables and the latent, unobserved variables that are theorized to explain these observed variables:

$$x = \Lambda_x \xi + \delta \qquad (1)$$

The associated covariance matrix of x, $\mathrm{E}(xx')$:

$$\mathrm{E}(\Lambda\xi + \delta)(\Lambda\xi + \delta)'$$

If $\mathrm{E}(\xi\xi') = \phi$ *and* $\mathrm{E}(\delta\delta') = \theta_\delta$ then:

$$\Sigma_{xx} = \Lambda_x \Phi \Lambda_x' + \theta_\delta \qquad (2)$$

consists of the partial regression coefficients (Λ) of the latent variables (ξ), the error variance (θ) and the factor covariance (Φ).

The goal of this stage of the analysis was simply to demonstrate that the first-order model for Statistics-Related Self-Efficacy with the sample of participants in this investigation converges directly on the items in the inventory. Equation (2) was used to generate congeneric models for the primary constructs of Statistics-Related Self-Efficacy. The second measurement model used later is the second order measurement model.

The Second-Order Structural Equation Model

A model, with the existence of lower-order factors and significant inter-correlations among the factors implies the existence of at least one second-order factor. Gorsuch (1983, p. 579) uses the following analogy to differentiate between a first-order and a second-order model:

> The first-order analysis is a close-up view that focuses on the details of the valleys and the peaks in mountains. The second-order analysis is like looking at the mountains at a greater distance, and yields a potentially different perspective on the mountains as constituents of a range.

In a broader theoretical framework, Gorsuch (1983, p. 240) distinguishes between the first-order factors (primary) and higher-order factors thus: "primary factors are concerned with narrow areas of generalization where the accuracy is great. The higher-order factor reduces accuracy for an increase in the breadth of generalization."

The confirmatory factor analyses of the second-order factor models were based on an extension of equation (1) and equation (2). Specifically, a second-order model involves re-specifying equation (1) and equation (2) as endogenous constructs rather than exogenous constructs. Endogenous constructs are different from exogenous constructs in that the former can be mediating constructs and pure dependent variables, whereas the latter are independent constructs.

The exogenous model is replaced by the following endogenous model:

$$y = \Lambda_y \eta + \varepsilon \qquad (3)$$

DOI: https://doi.org/10.7275/0nem-w659

in which the associated covariance matrix of y, $\mathrm{E}(yy')$, consists of the partial regression coefficients (Λ) of the latent variables (η), the error variance (ε) and the factor covariance (ψ):

$$\mathrm{E}(yy') \textit{ or } \mathrm{E}(\Lambda\eta+\varepsilon)(\Lambda\eta+\varepsilon)'$$

If $E(\eta\eta') = \psi$ and $E(\varepsilon\varepsilon') = \theta_\varepsilon$ then:

$$\Sigma_{yy} = \Lambda_y \Psi \Lambda_y' + \Theta_\varepsilon \quad (4)$$

The resulting model, in which the exogenous model (2) is replaced by the endogenous model (4), is then transformed into a second-order structured-mean model (Benson & Bandalos, 1992). Specifically, if ξ is the second-order factor, and η is the set of lower-order factors, equation (5) summarizes their causal relationship, Γ, between ξ and η, and ζ the residual associate with η:

$$\eta = \Gamma\xi + \zeta \quad (5),$$

with ζ, the residual variance associated with η. Combining equation (5) into equation (3) yields:

$$y = \Lambda_y(\Gamma\xi + \zeta) + \varepsilon \quad (6)$$

After taking the appropriate expectation of (6), the covariance of y for the second-order factor model is:

$$\Sigma_{yy} = \Lambda_y[\Gamma\Phi\Gamma' + \Psi]\Lambda_y'' + \Theta_\varepsilon \quad (7)$$

Equation (7) decomposes this covariance, to extract second order factor model, Λ_y, are the factor loadings, Γ the second order loadings, Φ the first order factor covariance, and θ_ε the associated residual matrix.

The analyses for this phase were conduced by specifying the LISREL parameters (7) for testing the first-order and second-order models simultaneously (Jöreskog & Sörbom, 2001, p. 204). The goal of these analyses was simply to demonstrate that Statistics-Related Anxiety maintained the hypothesized second-order factor structure by convergence on the items in each inventory.

Assessing Model Fit

A combination of criteria is utilized to assess the fit of the data to each model. Other indices are used to compare full models to reduced models. Satorra-Bentler's scaled corrections, SB χ^2 are used rather than the commonly used normal theory χ^2 because of the high level of kurtosis that is associated with items (Schermelleh-Engel, Moosbrugger, & Müller, 2003). The Satorra-Bentler scaled χ^2 (SB_{χ^2}) test adjusts the maximum likelihood estimators downward by a constant value which reflects the degree of the observed kurtosis, in an effort to minimize the effects of non-normality (Kline, 1998, p. 210). Data that is skewed or kurtotic can be problematic with the maximum likelihood estimation procedures.

However, because of the large sample size (n=238) in the present study, all the associated p values associated with the computed χ^2 did not exceed 0.05. Although χ^2 test statistic is the most commonly cited fit index, there are several problems with it. According to Schermelleh-Engel et al. (2003) and Satorra and Bentler (1999), the χ^2 test statistic is problematic when used with data that is not multivariate normal, it is extremely sensitive to sample size, and the χ^2 test statistic value decreases as model complexity increases; therefore, it is also affected by the number of parameters in the model.

Model fit is also evaluated by a combination of two relative fit indices: the Comparative Fit Index (CFI, Bentler, 2007) and the Root Mean Square Error of Approximation (RMSEA), which, according to Browne & Cudeck (1993, pp. 137-138) are designed to address the following issue: How well would the model, with unknown but optimally chosen parameter values fit the population covariance matrix if it were available? RMSEA demonstrates optimal fit with a value below 0.05, and a reasonably good fit with values at or below 0.08. CFI is non-stochastic, with p values greater than 0.95 indicating good fit. According to Fan, Thompson & Wang (1999) these indices have been shown to demonstrate very little random variation due to sample size, number of parameters, model misspecification, or method of estimation.

Brown and Cudeck (1993) also propose the consideration of a complementary question regarding model fit that focuses on the overall error in the model: How well can the model with parameter values determined from the available sample fit the population covariance matrix? They propose that cross-validation should be considered, and suggest that the expected value of the cross-validation index (ECVI) can be used to estimate the expected value of the cross-validation index based on the available sample. The smaller the ECVI, the smaller the discrepancy, and therefore, the better the model fit

Finally, Consistent Akaike Information Criterion (CAIC) is used to assess model fit in light of change in model complexity. The use of this information fit index defines a selection criterion that makes an appropriate adjustment to its goodness of fit by penalizing for model complexity (Myung, 2000, p.196). Lower CAIC values indicate better fit.

Jackknifing Procedure

The current investigation demonstrates a jackknife approach, in which individual items are removed after the full model is estimated. The jackknife procedure, similar to the procedure presented by Rensvold and Cheung (1999), is applied to item reduction with the following procedure:

Step 1: The fit statistics are calculated for the full data set with all items;

Step 2: Re-estimate the model, K number of times, with each estimate based on the full model minus one of the items, with a different item removed for each re-estimate.;

Step 3: Rank the resulting models and determine which model has the best fit relative to the original full-item model, based on CFI and RMSEA values; and

Step 4: With the best fitting model identified in Step 3, repeat the procedure starting with Step 2, this time re-estimating the one-item reduced model.

Step 5: Continue this model re-estimating and item-removal process until the following conditions are met:

(i) Variables were removed from the models as long as the original primary factor model was correlated with the reduced model at a level of $r \geq .95$, as recommended by Newcomb et al. (1988) and Byrne (1989).

(ii) Items were removed as long as each original factor continued to explain at least three observed variables (Bagozzi, 1980; Sluis et al., 2005).

(iii) Items were removed as long as the structural integrity of the model was not violated (Bollen, 1989);

(iv) And the resulting reduced model demonstrated good fit (Bollen, 1989).

While these procedures are computationally intensive, an automated application was developed with FORTRAN for use in the present study.

Bootstrap Confidence Band

For the items that the full model and the reduced model had in common, bootstrap confidence intervals were also computed as another means of examining the reliability of the inventories with this sample of participants. With this procedure, the bootstrap sample data was based on a sample 1 ½ times the original sample size (n = 357), with replacement, and replicated 2000 times. The purpose of this procedure was to produce 95% confidence bands in an effort to assess the precision of the parameters. Specifically, the tighter the confidence bands, the higher the level of precision associated with the coefficient (Sturgis, 2005). Narrow bandwidths for most of the items suggest a high level of reliability in the original model data (Sturgis, 2005; Wood, 2005).

Methods

Sample

The participants were 238 graduate level students enrolled in statistics courses offered in the departments of biological science, education, geography, and psychology at Kent State University. The students completed paper and pencil questionnaires examining self-beliefs and emotions about their required statistics coursework.

Instrumentation

The Current Statistics Self-Efficacy (CSSE) inventory is an instrument developed by Finney and

Schraw (2003) to assess the one-dimensional construct of self-efficacy. With this instrument, respondents are asked to rate their current belief in their ability to complete 14 specific tasks related to statistics using a 1 (no confidence at all) to 5 (complete confidence) response scale.

The specific items for this inventory are provided in Appendix A.

The Statistics Anxiety Rating Scale (STARS-1) is an instrument developed from the original STARS inventory by Cruise, Cash, & Bolton (1985). The original STARS inventory was created to assess two different factors related to statistics anxiety – specifically, anxiety and attitudes about statistics. The STARS-1 is comprised of the first twenty-three items of the STARS (Cruise et al., 1985) as a measure of statistics anxiety across three subscales: (1) anxiety related to interpretation; (2) statistics class and test anxiety; and (3) anxiety about asking for assistance. The items for this inventory are provided in Appendix B.

Reliability analyses were conducted using SPSS 12.0.1 (SPSS, 2003) in order to assess the consistency of participant responses on the scales. A Cronbach's Alpha (Cronbach, 1951) was calculated for the data collected with each instrument (CSSE and STARS) in an effort to analyze the internal consistency of items in each scale. These analyses, conducted on the ordinal responses, revealed acceptably high levels of reliability (Thompson, 2003, p. 256) for each instrument, with an a = .917 on the 14 items of the CSSE inventory, and an a = .917 on the 23 items of the STARS-1 inventory. A breakdown of the reliability analyses for each construct is presented in Table 1.

Table 1: *Cronbach's Alpha for Primary Factors*

Construct	Sub-Construct	Number of Items	Cronbach's α
Self-Efficacy	One-Dimensional	14	.917
Anxiety	Interpretation	11	.842
	Class/Test	8	.889
	Assistance	4	.776

A full description of the sample and the procedures is provided in Larwin (2007).

Data Preparation

Once data were collected, a number of procedures were used to prepare the data for subsequent analyses. First, data were examined for missing values. A total of sixteen item-responses were incomplete. Since there was no pattern to the missing responses, multiple imputation procedures, generated through the Linear Structural Relations program (LISREL® 8.8, 2006), were used to complete the sixteen missing responses.

Multiple imputation is one of many methods available for dealing with missing data (Fox-Wasylyshyn & El-Masri, 2005; McCleary, 2002). Multiple imputation was implemented in the present study because it is considered by many researchers to be the superior approach to dealing with missing data (e.g., Allison, 2000; Fishman & Cummings, 2003; King, Honaker, Joseph, & Scheve, 2001; Rubin, 1987; Schafer & Olsen, 1998), and unlike other methods, multiple imputation has been found to be robust to model violations (Allison, 2000; King, et al., 2001). Multiple imputation is accomplished through several stages of data analyses in which data from complete cases is used to predict the value of the missing item.

Some of the survey responses were recoded in an effort to have all responses across the two inventories coded in such a way that the theoretically least desirable responses had the lowest values and the most desirable responses had the highest values. Specifically, items were coded so that item-responses indicating highest-level of anxiety, lowest-level of self-efficacy and poorest attitudes were recoded as having a value of zero; responses indicating lowest-levels of anxiety, highest-levels of self-efficacy, and most positive attitudes were recoded as a value of four (4). The methods of analysis in the next section are confirmatory factor analysis and jackknife item-reduction procedure.

Results

Part One: The Congeneric Model of Statistics-Related Self-Efficacy

For Part One of the present investigation, data from Finney & Schraw's (2003) Current Statistics Self-Efficacy inventory (CSSE) that was employed to measure participants' Statistics-Related Self-Efficacy was used to examine the viability of using jackknife procedures for item reduction. This instruments' one-dimensional structure was ideally suited for the first part of the current investigation.

In an effort to demonstrate the construct validity of the instrument with the present sample of participants, it was necessary to verify the factor structure of the inventory through confirmatory factor analysis (CFA), as well as demonstrate that the 14 items load significantly on the factor. Specifically, the proposed covariance matrix for the primary factor (Σ), which has been supported by prior research (Finney & Schraw, 2003) was tested against the sample covariance matrix (*S*) (Bollen, 1989). In order to accomplish this, each item was constrained to load on only one first-order factor, hence making this a confirmatory factor model (Lord &Novick, 1968). Equation (2) is operationalized with the Statistics-Related Self-Efficacy construct (Figure 1)

$$Self-Efficacy\,\Lambda_y = \begin{bmatrix} ITEM & Self-Efficacy \\ EF1 & \lambda_{1,1} \\ EF2 & \lambda_{2,1} \\ EF3 & \lambda_{3,1} \\ EF4 & \lambda_{3,1} \\ EF5 & \lambda_{5,1} \\ EF6 & \lambda_{6,1} \\ EF7 & \lambda_{7,1} \\ EF8 & \lambda_{8,1} \\ EF9 & \lambda_{9,1} \\ EF10 & \lambda_{10,1} \\ EF11 & \lambda_{11,1} \\ EF12 & \lambda_{12,1} \\ EF13 & \lambda_{13,1} \\ EF14 & \lambda_{14,1} \end{bmatrix},$$

and

$$\Theta_{\delta} = \delta_{1,1}, \delta_{2,2}, \delta_{3,3}, \ldots, \delta_{14,14} \quad and \quad \Phi = \mathrm{I}$$

Figure 1. Operationalization of the Self-Efficacy Construct

The results of the CFA indicate that the Statistics-Related Self-Efficacy model demonstrated good fit (χ^2 = 598.64, CFI = .948, RMSEA = .127, ECVI = 1.808, CAIC = 553.82). The RMSEA values for this model are higher than what is considered acceptable. However, Schermelleh-Engel et al. (2003) indicate that the RMSEA calculations are sensitive to the number of variables in a congeneric model.

Jackknife Application to CSSE

Once the original CSSE model was confirmed through CFA, it was re-estimated as a new model that systematically reduced the number of items in the model by one. After each iteration, the fit statistics were recorded following the steps outlined in Jackknife Procedures (above). For this 14 item model, this process required a total of 881 separate LISREL runs. The items from the item-deletion procedure, which created the best fitting model based on the CFI and RMSEA estimates, were manually deleted from the subsequent runs (with K – 1 items) of the model. The Statistics-Related Self-Efficacy Model was reduced by a total of five items; a reduction of 35.7%. The deleted items are presented in Table 2.

Table 2: *Jackknife Item-Elimination Results for CSSE*

Jackknife Subset (item deleted)	χ^2	*df*	CFI	RMSEA	ECVI	CAIC
5	523.04	65	.951	.127	2.686	752.80
10	398.25	54	.958	.118	2.040	590.90
6	314.93	44	.967	.109	1.547	465.05
1	243.02	35	.965	.111	1.291	395.38
13	183.76	27	.968	.109	.992	315.49

Once items were removed, the items in the reduced factor model were correlated with the same items in the original larger factor model. This resulted in an r = .966, satisfying the recommended r ≥ .95 guidelines of Newcomb & Bentler (1988) and Byrne (1989). The primary factor continued to converge on at least three observed variables (Bagozzi, 1980; Sluis et al., 2005), and the original model integrity was maintained, with the reduced model demonstrating good fit. As revealed in Table 3, the final model of 9 items demonstrated an improved fit relative to the full model of 14 items.

Table 3: *Self-Efficacy Change in Model Fit*

	χ^2	*df*	CFI	RMSEA	EVCI	CAIC
All Items (14 Items)	598.65	77	.948	.127	1.808	553.82
Reduced Model (9 Items)	183.76	27	.968	.109	.923	315.49
Δ Self-Efficacy	***414.88***	***50***	***.020***	***.018***	***.885***	***238.33***

The deleted and retained items from the CSSE are presented in Tables 4a and 4b.

Table 4a: *Self-Efficacy Items Retained with Model Reductions*

Retained Item	Retained Item Content
EF2	Interpret the probability value from a statistical procedure.
EF3	Identify if a distribution is skewed when given the values of three measures of central tendency.
EF4	Select the correct statistical procedure to be used to answer a research question.
EF7	Explain what the value of the standard deviation means in terms of the variable being measured.
EF8	Distinguish between a Type 1 error and a Type 2 error in hypothesis testing.
EF9	Explain what the numeric value of the standard error is measuring.
EF11	Distinguish between the information given by the three measures of central tendency.
EF12	Distinguish between a population parameter and a sample statistic.
EF14	Explain the difference between a sampling distribution and a population distribution.

Table 4b: *Self-Efficacy Items Deleted with Model Reductions*

Deleted Item	Deleted Item Content
EF1	Identify a scale of measurement for a variable.
EF5	Interpret the results of a statistical procedure in terms of the research question.
EF6	Identify the factors that influence power.
EF10	Distinguish between the objectives of descriptive versus inferential statistical procedures.
EF13	Identify when the mean, median, and mode should be used as a measure of central tendency.

Table 5 demonstrates that the factor loadings for the items that are common to both the 9-Item and 14-Item models, indicate minimal change in the loadings as a result of the item-reduction. The differences in factor loadings range from -0.067 to 0.084, with a mean difference of M = -0.006, SD = 0.054 indicating a minor decrease in loadings after the item deletion. Negative values indicate that the influence of the first-order factors on the observed variables dropped as a result of the item elimination. The largest drop in loadings was with item EF2 ($\delta\Lambda$ = -0.067).

Table 5: *Comparisons of First-Order factor loadings for Statistics-Related Self-Efficacy (Λ_y)*

Item	Reduced Model	Full Model Parameter	$\delta\Lambda$
EF2	**.998***	**1.065***	**-.067**
EF3	1.243*	1.177*	.066
EF4	.656*	.704*	-.048
EF7	1.056*	1.065*	-.009
EF8	1.271*	1.302*	-.031
EF9	.924*	.970*	-.046
EF11	1.677*	1.593*	.084
EF12	1.347*	1.312*	.035
EF14	1.106*	1.140*	-.034

Note: **p* <.05, and bolded item was fixed to 1.0.

Bootstrapped confidence intervals were also computed as another means of examining the reliability of the inventories with this sample of participants. (Wood, 2005) The results are reported in Table 6. The narrow bandwidths suggest a high level of reliability in the original model data (Sturgis, 2006). The bootstrap confidence bands for the items in the reduced model of Statistics-Related Self-Efficacy ranged from 0.017 to 0.056 with a mean of M = 0.039, SD = .014; the average change (δ) in bandwidths from the original model to the reduced model was a minimal increase (M = .016, SD = .013). Negative values indicate a decrease in bandwidth from the reduced model to the full model, indicating an increase in precision as a result of the item reduction.

Table 6: *Retained Statistics-Related Self-Efficacy Bootstrap Confidence Intervals* (*N* = 2000)

Item	Coefficient	SE	Lower Bound	Upper Bound	Reduced Model	Full Model	δ Bandwidth
EF2	0.849	.007	0.835	0.863	.028	.018	.010
EF3	1.103	.011	1.082	1.124	.043	.024	.019
EF4	0.053	.004	0.045	0.062	.017	.023	-.006
EF7	0.889	.010	0.870	0.908	.037	.024	.013
EF8	1.117	.013	1.091	1.143	.052	.023	.029
EF9	0.779	.006	0.768	0.790	.022	.023	-.001
EF11	1.432	.014	1.404	1.460	.056	.024	.032
EF12	1.180	.014	1.154	1.207	.053	.023	.030
EF14	0.959	.012	0.936	0.982	.041	.023	.024

Overall, the reliability of the reduced 9-item model, relative to the 14-item model, as measured with the Squared Multiple Correlation coefficient, also demonstrated an improvement in the models reliability, from 0.533 for the full model to 0.578 for the reduced model. This suggests that item reduction was successfully accomplished using suggested guidelines without compromising the measurement integrity of the original model

Part Two: The Multi-Dimensional Model of Statistics-Related Anxiety

Part Two of this investigation used data collected with the Statistics Anxiety Rating Scale-1 (STARS-1, Cruise, Cash, & Bolton, 1985). The STARS-1 is theoretically conceptualized as a second order factor structure that is made up of three primary factors and 23 observed items. This adds an additional level of complexity to the process of item reduction in the current investigation. Specifically, the presence of a second order factor, along with the three primary factors, potentially complicates the jackknifing procedures ability to comply with the second and third decision rules guiding the item-reduction process (i.e., "Items were removed as long as each original factor continued to explain at least three observed variables," and "Items were removed as long as the structural integrity of the model was not violated"). As illustrated in the conceptual diagram in Figure 2, the second-order factor, Anxiety, has three first-order anxiety-related factors: Interpretative Anxiety, Class/Test Anxiety, and Assistance Anxiety.

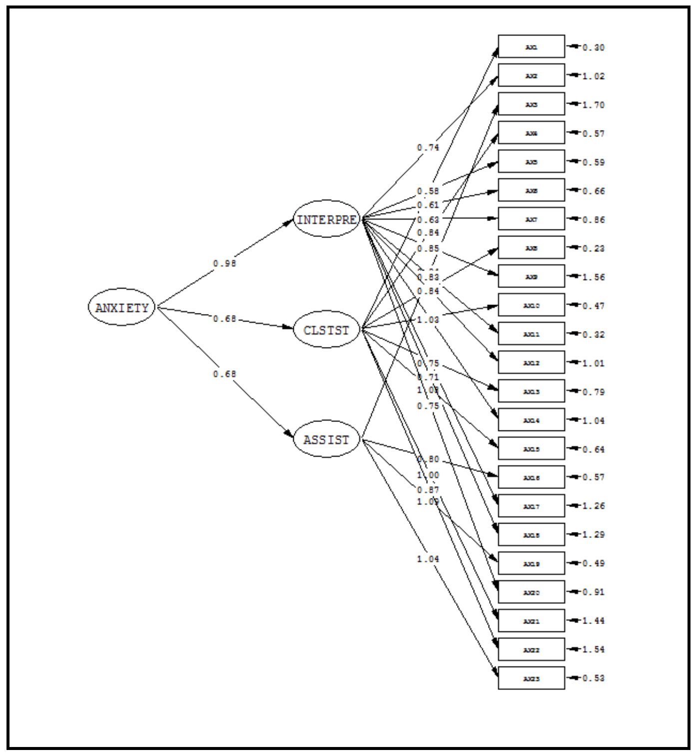

Figure 2. Three-factor CFA for Statistics-Related Anxiety- STARS

CFA Establishing Construct Validity of Statistics-Related Anxiety

In an effort to demonstrate the construct validity of the instrument, it was necessary to verify the factor structure of the inventory through confirmatory factor analysis (CFA), as well as demonstrate that the items identified by the factor were stable. Specifically, the population covariance matrix was tested against the sample covariance matrices (***S***) (Bollen, 1989). In order to accomplish this, one item was constrained to load on only one first-order factor, and one item on each factor was constrained to be equal to one in order to establish the metric for the factor (Benson & Bandalos, 1992).

Equation (7) as operationalized, with the Statistics-Related Anxiety construct is shown in Figure 3.

$$
\text{Anxiety } \Lambda_y = \begin{bmatrix}
ITEM & Interpretation & Class/Test & Assist \\
AX1 & 0 & 1 & 0 \\
AX2 & 1 & 0 & 0 \\
AX3 & 0 & 0 & 1 \\
AX4 & 0 & \lambda_{4,7} & 0 \\
AX5 & \lambda_{5,6} & 0 & 0 \\
AX6 & \lambda_{6,6} & 0 & 0 \\
AX7 & \lambda_{7,6} & 0 & 0 \\
AX8 & 0 & \lambda_{8,7} & 0 \\
AX9 & \lambda_{9,6} & 0 & 0 \\
AX10 & 0 & \lambda_{10,7} & 0 \\
AX11 & \lambda_{11,6} & 0 & 0 \\
AX12 & \lambda_{12,6} & 0 & 0 \\
AX13 & 0 & \lambda_{13,7} & 0 \\
AX14 & \lambda_{14,6} & 0 & 0 \\
AX15 & 0 & \lambda_{15,7} & 0 \\
AX16 & 0 & 0 & \lambda_{16,8} \\
AX17 & \lambda_{17,6} & 0 & 0 \\
AX18 & \lambda_{18,6} & 0 & 0 \\
AX19 & 0 & 0 & \lambda_{19,8} \\
AX20 & \lambda_{20,6} & 0 & 0 \\
AX21 & 0 & \lambda_{21,7} & 0 \\
AX22 & 0 & \lambda_{22,7} & 0 \\
AX23 & 0 & 0 & \lambda_{23,8}
\end{bmatrix}
$$

Figure 3. Anxiety Lambda Matrix

As can be seen in Figure 3, Equation (7) is expanded from the first-order factor model by incorporating two additional indices. The first-order model was specified as

$$\Theta 1st = [\ \Lambda x, \Psi, \Theta\delta\ ],$$

whereas for the second-order model, Θ becomes

$$\Theta = [\Lambda y, \Psi, \Theta\varepsilon\ , \Gamma, \Phi].$$

The additional matrices in the second-order model are operationalized such that $\Phi = I$ and Γ is presented in Figure 4.

$$
\Gamma_{AX} = \begin{bmatrix}
\eta & Anxiety \\
Interpretation & 1 \\
Class/Test & \gamma_2 \\
Assistance & \gamma_3
\end{bmatrix}
$$

Figure 4. Second-order covariance matrix of Anxiety

CFA results supported the strict confirmatory factor structure of each of the primary factors in the present study. Basically, when tested, each of the theorized models was found to have an acceptable level of fit to the sample of data without any further modifications. The results of the CFA indicate that Interpretative Anxiety explained eleven of the STARS-1 items; Class/Test Anxiety explained eight of the STARS-1 items; and Assistance Anxiety explained four of the STARS-1 items. The Statistics-Related Anxiety model demonstrated good fit, ($\chi^2$3 = 1235.91, CFI = 0.946, NNFI = 0.923, RMSEA = 0.095).

Jackknife Application to STARS

The original STARS-1 model was estimated, and then re-estimated following the item deletion steps presented in the *Jackknife Procedures* (above). The model was estimated for each subset of items, and the fit statistics were recorded. For this 23 item model, this process required a total of 3,309 separate LISREL runs. The items from the item-deletion process which created the best fitting model, based on the CFI and RMSEA estimates were deleted from the subsequent runs (with K – 1 items) of the model. As a result of these procedures, The Statistics-Related Anxiety Model was reduced by a total of nine items; a reduction of 39.1%. The deleted items and the fit indices prior to their deletion are presented in Table 7.

Table 7: *Jackknife Item-Elimination Results for STARS-1*

Jackknife Subset (item deleted)	χ^2	*df*	CFI	RMSEA	ECVI	CAIC
23	977.43	206	.952	.096	4.963	1386.36
17	949.95	186	.949	.095	4.729	1322.06
4	904.49	166	.969	.074	4.514	1266.66
20	776.01	148	.976	.066	3.914	1115.45
18	907.99	131	.938	.108	4.926	1346.26
6	555.64	115	.961	.088	2.816	837.25
14	520.41	100	.955	.094	2.670	793.90
8	393.08	87	.096	.083	1.930	605.00
10	308.52	74	.968	.076	1.523	499.58

A list of each deleted and retained item is presented in Tables 8 and 9.

Table 8: *Anxiety Items Deleted by Model Reductions*

Item	Deleted Item Content
AX4	Doing the homework for a statistics course.
AX6	Reading a journal article that includes some statistical analysis.
AX8	Dong the final examination in a statistics class.
AX10	Walking into the classroom to take a statistics test.
AX14	Figuring out whether to reject or retain the null hypothesis.
AX17	Trying to understand the odds in a lottery.
AX18	Seeing a student pouring over the computer printouts related to his/her research.
AX20	Trying to understand the statistical analysis described in an abstract of a journal article.
AX23	Asking a fellow student for help in understanding a printout .

Once the items were removed, the original primary factor model was correlated with the reduced model, r = .959, satisfying Newcomb & Bentler's (1988) and Byrne's (1989) recommended r ≥ .95 guidelines; and as illustrated in Figure 4 and Table 10, the three primary factors continued to be explained by at least three observed variables (Bagozzi, 1980; Sluis et al., 2005).

Prior to the full model and reduced model analyses, the loading for one of each of the three first-order factors, explained by Anxiety, was set to one, as an anchor for the purpose of model identification. Specifically, item AX2 of Interpretation, item AX1 of Class/Test, and item AX3 of 'Assist' were set to 1.0. All factor loadings and reliabilities remain significant after the jackknife procedure was completed.

Table 9: *Anxiety Items Retained after Reductions*

Item	Retained Item Content
AX1	Studying for an examination in a statistics course.
AX2	Interpreting the meaning of a table in a journal article.
AX3	Going to ask my statistics teacher for individual help with material I am having difficulty understanding.
AX5	Making an objective decision based on empirical data.
AX7	Trying to decide which analysis is appropriate for your research project.
AX9	Reading an advertisement for an automobile which includes figures on gas mileage, compliance with population regulations, etc.
AX11	Interpreting the meaning of a probability value once I have found it.
AX12	Arranging to have a body of data put into the computer.
AX13	Finding another student in class got a different answer than you did to a statistical problem.
AX15	Waking up the morning on the day of a statistics test.
AX16	Asking one of your professors for help in understanding a printout.
AX19	Asking someone in the computer center for help in understanding a printout.
AX21	Enrolling in a statistics course.
AX22	Going over the final examination in statistics after it has been graded.

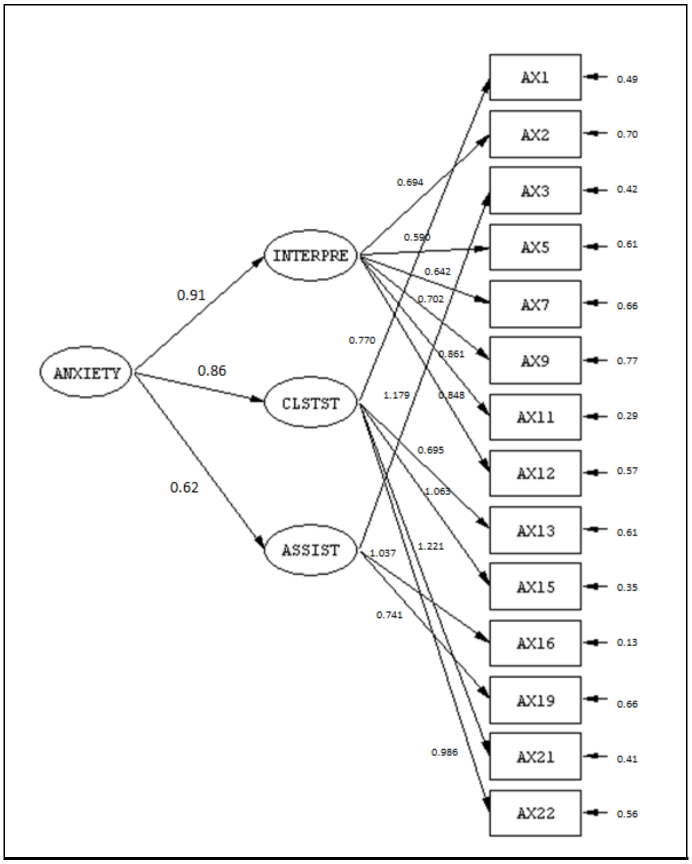

Figure 4. Fourteen Remaining Items of Statistics-Related Anxiety Model

Table 10: *First-order factor loadings for Statistics-Related Anxiety (Λ_y)*

1st Order Factor	Reduced Model Parameter Estimate	Full Model Parameter Estimate	$\delta\Lambda$
INTERPRET			
AX2	0.694*	0.752*	-.058
AX5	0.590*	0.597*	-.007
AX7	0.642*	0.638*	.004
AX9	0.702*	0.820*	-.118
AX11	0.861*	0.849*	.012
AX12	0.848*	0.828*	.020
CLASS/TEST			
AX1	0.770*	0.912*	-.142
AX13	0.695*	0.716*	-.021
AX15	1.063*	1.092*	-.029
AX21	1.221*	1.058*	.163
AX22	0.986*	0.092*	.094
ASSISTANCE			
AX3	1.179*	1.032*	.147
AX16	1.037*	0.917*	.120
AX19	0.741*	0.956*	-.215

Note: *p <.05 and bolded items were fixed to 1.0

As indicated in Table 10, the differences in factor loadings range from -0.215 to 0.163, with a mean difference of M = 0.003, SD = 0.109, indicating an overall increase in effect after the deletion of nine items. The largest drop in covariance was with item AX19 ($\delta\Lambda$ = -0.215) and item AX1 ($\delta\Lambda$ = -0.142). As stated above, negative values indicate that the influence of the first-order factors on the observed variables dropped as a result of the item elimination.

Additionally, it was important that not only the original model integrity was maintained, but that the reduced model demonstrated good fit as well. As revealed in Table 11, the final reduced model demonstrated an improved fit relative to the full model. Each of the respective fit indices demonstrated improvement as a result of the item deletion.

Table 11 : *Goodness-of-Fit Indices for Primary Factors after Item Elimination*

Construct	$_{SB}\chi^2$	*df*	CFI	RMSEA	ECVI	CAIC
Anxiety 23 Items	1149.88	227	.946	.0947	5.667	1562.19
Anxiety 14 Items	308.52	74	.968	.0763	1.523	499.51
Δ Anxiety	***841.36***	***153***	***.022***	***.0184***	***4.144***	***1062.68***

Because anxiety was the second-order factor that explained the variance associated with the first-order factors (Interpretation, Class/Test Anxiety, and Assistance) it is also important to examine how the model works at the different levels of factor structures. The factor loadings and reliabilities for the first- and second-order constructs are present in Table 12 and Table 13.

Table 12: *Squared-Multiple Correlations (R^2) of First-Order and Second-Order Factors (β)*

2nd Order	1st Order	Reduced Model	Full Model	Δ R^2
Anxiety		.847*	.834*	-.013
	Interpretation	.735*	.665*	-.070
	Class/Test	.836*	.782*	-.054
	Assistance	.332*	.404*	-.172

Note: *p <.05 and bolded items were fixed to 1.0; Reliability is multiple R^2 value.

Table 13: *Comparison of Factor Loadings of Second Order Factors (β)*

2nd Order	1st Order	Reduced-Item Parameter Estimate	Full Model Parameter Estimate	Δβ
Anxiety	Interpretation	.857*	.815*	.042
	Class/Test	.929*	.844*	.045
	Assistance	.576*	.636	.060

Note: *p <.05

All of the factor loadings (β) and reliabilities between the second-order factor and each of the first-order factors were found to be significant and improved after the jackknife procedure. This indicates that the second-order construct adequately indicates the variance shared by the first-order constructs. Although the reliability of the Assistance factor (R^2 = 0.332) is weak, this was not a concern as these items did not affect the overall good fit of the model (Brunner and Süb, 2005, p. 237).

As indicated in Table 14, the confidence bands for the Reduced-Item model of Statistics-Related Anxiety ranged from 0.016 to 0.082, with a mean change in bandwidth of M=0.031, SD=0.017. The values for the δ bandwidth are relatively close to zero. This is further indication that the Reduced-Item model adequately reflects the Full Model.

Table 14: *Retained Items Statistics-Related Anxiety Bootstrap Confidence Intervals (N = 2000)*

Item	Coefficient	SE	Lower Bound	Upper Bound	Reduced Model	Full Model	δ Bandwidth
AX2	.554	.007	.540	.568	.028	.022	.006
AX3	.675	.008	.659	.691	.033	.021	.012
AX5	.489	.005	.480	.498	.019	.024	-.005
AX7	.501	.004	.493	.509	.016	.016	.000
AX9	.438	.021	.397	.479	.082	.047	.035
AX11	.685	.006	.673	.697	.025	.024	.001
AX12	.594	.007	.580	.609	.029	.024	.005
AX13	.607	.006	.596	.618	.022	.020	.002
AX15	.843	.007	.830	.857	.027	.021	.006
AX16	.514	.005	.505	.523	.018	.026	-.008
AX19	.542	.012	.518	.566	.048	.032	.016
AX21	.938	.009	.921	.955	.034	.024	.010
AX22	.768	.008	.752	.784	.032	.022	.010

One potential concern with the present procedure is that the change in the models' goodness-of-fit index exceeded Cheung and Rensvold's (2002) recommended level of ΔCFI <0.01 for measurement invariance. However, the changes in goodness-of-fit, as a measure of invariance, are sensitive to model complexity, in the form of numbers of items, numbers of factors, and the ratio of the two (Wu, Li, & Zumbo, 2007, p. 5). In light of this, Wu et al. suggest that less stringent cut-offs are appropriate when considering measurement invariance where a change in model complexity is an issue, as is the case with the present investigation. Because Wu et al. do not provide detailed guidelines, a test of whether a significant change had occurred in the model's complexity was tested. A ratio t test was conducted in an effort to determine whether a significant change in the complexity levels of the primary factor, as indicated by the change in the ratio of number of factors to number of items (Wu et al., 2007) was present. If a significant change was revealed, it would be reasonable to consider the change in goodness-of-fit, from original models to reduced models, measurement invariant. Table 15 presents the results of the ratio t-test assessing the change in model complexity from the Full Item model to the Reduced Item model.

Table 15: *Ratio t Test of Model Complexity*

Factor	Full Model Ratios	Reduced Model Ratios	Ratio Change	Ratio t-test
SELF-EFFICACY	.071	.111	.040	.19*
ANXIETY				
Interpretation	.091	.200	.109	.34*
Class/Test	.125	.167	.042	.13*
Assistance	.250	.333	.083	.12*

Note: * $p < 0.05$

The ratio t-test was used to analyze the differences in model complexity because model complexity is expressed as a ratio (Myers & Wells, 2003). These values were computed with the Ratio t test which computes the logarithm of the ratios:

$$\log(\text{65-Item Ratio}/\text{40-Item Ratio}) = \log(65\,\textit{Item ratio}) - \log(40\,\textit{Item ratio})$$

For this analysis, no change in the ratio, indicating no change in model complexity, is equal to zero, the logarithm of 1.0. As indicated in Table 16, all ratios indicated a decrease in model complexity, at the level of each first-order factor, that was statistically significant.

Discussion

For the present investigation, each primary factor model had an average item reduction of 38% of its items after the item elimination process. And after item elimination, the original single-construct models correlated with their respective reduced models at a level of $r \geq .95$, as recommended by Newcomb & Bentler (1988) and Byrne (1989), indicating that the fundamental nature of the model has not been changed. The full model was found to be appropriately sufficiently correlated with the reduced model, according to these guidelines , for both Statistics-Related Self-Efficacy ($r = .966$), and Statistics-Related Anxiety ($r = .959$). Additionally, items were removed in the jackknife procedure as long as each original factor continued to explain at least three observed variables (cf. Bagozzi, 1980; Sluis et al., 2005).

While approaches to item reduction have been proposed by other researchers (e.g., Benson & Bandalos, 1992; Clark & Goldsmith, 2006; Salzberger, 2006), these approaches have failed to produce resulting models that were both measurement invariant and structurally invariant. In addition, these approaches fail to offer clear guidelines as to what is an acceptable level of model reduction before the construct is considered to be successfully condensed. While the novel approach presented here is still in its infancy, it provides a clear method of item reduction that seems to result in more parsimonious models that are measurement invariant, structurally invariant, and demonstrate better fit by maintaining only those items that are truly working well in the model. The present approach is superior to earlier research in that it addresses the primary point of concern in SEM, that being model fit. This investigation incorporates five different fit indices, as well as bootstrapping analyses, in an effort to demonstrate the influence of each item's elimination on the fit statistics ($SB\chi^2$, CFI, RMSEA, ECVI, and CAIC respectively) by determining how much each statistic would be improved if a specific item was eliminated from the data set. This provides an unambiguous indication of individual item influence in the context of a specific structural equation model.

Although the jackknife procedure presented with these two models has worked when applied to a variety of structural models, the magnitude of these effects on models with different specifications is unknown. The two models presented here, and the other known models that have been tested, demonstrated good stability and improved fit, when this procedure is implemented according to the outline decisions rules stated above. The procedure described here is more fully discussed in Larwin (2007). A computer program, for performing this procedure within the FORTRAN format, is available from the author upon request.

REFERENCES

Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Sociological Methods and Research, 28*(3), 301-309.

Bagozzi, R. P. (1980). *Causal Models in Marketing.* New York: John Wiley & Sons, Inc.

Benson, J., & Bandalos, D. (1989). Structural model of statistical test anxiety. In R.Schwarwer, H.M. Vander Ploeg, & C.D.Spielberger (Eds) *Advances in Test Anxiety Research,*(Vol. 5, pp.137-154) Hillsdale, NJ: Erlbaum.

Bentler, P. M. (2007). *EQS Structural Equation Program.* Version 6, Encino, CA: Multivariate Software, Inc.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & S. Long, (Eds), *Testing Structural Equation Models.* Newbury Park CA: SAGE.

Brunner, M., & Süb, H. M. (2005). Analyzing the reliability of multidimensional measures: An example from intelligence research. *Educational and Psychological Measurement, 65*(2), 227-240.

Byrne, B. M. (1989). *A Primer of LISREL: Basic Applications and Programming for Confirmatory Factor Analytic Models.* New York: Springer-Verlag.

Chen, F., Bollen, K. A., Paxton, P., Curran, P. and Kirby, J. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies, *Sociological Methods and Research*, 29, 468–508.

Cheung, G.W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

Clark, R. A., & Goldsmith, R. E. (2006). Global innovativeness and consumer susceptibility to interpersonal influence. *Journal of Marketing Theory and Practice, 14*(4), 275-285.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Cruise, R., Cash, R., & Bolton, D. (1985). *Development and validation of an instrument to measure statistics anxiety.* Proceedings of the American Statistical Association. Alexandra, VA.

Dillon, W. R., Kumar, A. and Mulani, N. (1987), Offending estimates in covariance structure analysis: Comments on the causes and solutions to Heywood cases, *Psychological Bulletin,* 101, 126–135.

Donders, J. (1997). A short form of the WISC-III for clinical use, *Psychological Assessment*, 9, 15-20.

Epstein, R. (1984). "The Principle of Parsimony and Some Applications in Psychology". *Journal of Mind Behavior, 5*, 119–130.

Fan, X., Thompson, B., &Wang, L. (1999). The effects of sample size, estimation methods, and model specification on SEM fit indices. *Structural Equation Modeling,* 6, 56-83.

Fishman, M., & Cummings, J.N. (2003). Multiple imputation for missing data: Making the most of what you know. *Organizational Research Methods, 6*(3), 282-308.

Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology, 28,* 161-186.

Fox-Wasylyshyn, S. M., & El-Masri, M. M.(2005). Handling missing data in self-report measures. *Research in Nursing and Health, 28*, 488-495.

Freedman, D. A. (1987). As others see us: A case study in path analysis, *Journal of Educational Statistics,* 12, 101-128.

Gorsuch, R. L. (1983). *Factor Analysis* (2nd Ed.) New Jersey :Lawrence Erlbaum Associates, Inc.

Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Marketing Science, 23*(4), 519–529.

Grossi, E., Groth, N., Mosconi, P., Cerutti, R., Pace, F., Compare, A. & Apolone, G. (2006). Development and validation of the short version of the

Psychological General Well-Being Index (PGWB-S), *Health and Quality of Life Outcomes*, *4*, 85-88

Hanson, F. (2003, December). *Computational Stochastic Control: Basic Foundations, Complexity and Techniques***.** Proceedings of 2003 Conference on Decision and Control, Invited Poster/Interactive Paper in a Control Education Session. Maui, Hawaii.

Hayduk, L. A. (1987). *Structural Equation Modeling with LISREL*. London: John Hopkins University Press.

Heywood, H. B. (1931). On finite sequences of real numbers, *Proceedings of the Royal Society of London, Series A, Containing Papers of a Mathematical and Physical Character*, 134, 486–501.

Jackson, C. J., A. Furnham, L. Forde and T. Cotter (2000). The structure of the Eysenck Personality Profiler, *British Journal of Psychology*, 91, 223-239.

Jöreskog, K.G., & Sörbom, D. (2001), *LISREL 8: User's Reference Guide*, Illinois: Scientific Software International, Inc.

Kaplan, D. (1994). Estimator Conditioning Diagnostics for Covariance Structure Models. *Sociological Methods & Research*, Vol. 23, No. 2, 200-229

King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review, 95*, 49-69.

Kline, R.B. (1998). *Principles and Practices of Structural Equation Modeling*. New York: Gilford Press.

Larwin, K. (2007). *A structural equation model examining students' prior mathematics/statistics experiences and self-perception regarding graduate level statistics coursework: A methodological investigation.* (Unpublished doctoral dissertation). Kent State University, Kent, Ohio.

Lim, S., & Melville, N.P. (n.d.) Robustness of Structural Equation Modeling to Distributional Misspecification: Empirical Evidence & Research Guidelines, Electronic copy available at: http://ssrn.com/abstract=1375251

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

MacCallum, R. C., & Austin, J.T. (2000). Application of structural equation modeling in psychological research, *Annual Review of Psychology*, 51, 201-226.

Malhotra, N. K., Peterson,M., & Kleiser, S.B. (1999). Marketing Research: A State-of-the-Art Review and Directions for the Twenty-First Century, *Journal of the Academy of Marketing Science*, 27, 160-183.

Marsh, H. W., Dowson, M., Pietsch, J. and Walker, R. (2004). "Why multicollinearity matters: a reexamination of relations between self-efficacy, self-concept, and achievement", Journal *of Educational Psychology*, 96, 518–522

Martens, M. P. & Haase, R.F. (2006). Advanced Applications of Structural Equation Modeling in Counseling Psychology Research, *The Counseling Psychologist*, 34 ,878-911.

McCleary, L. (2002). Using multiple imputation for analysis of incomplete data in clinical research. *Nurs. Res. 51,* 339–343.

Mela, C. F. and Kopalle, P.K. (2002), The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlations, *Applied Economics*, 34, 667-677

Myers, J. L., & Well, A.D. (2003). *Research design and statistical analysis* (2nd ed.). Mahwah, NJ, Lawrence Erlbaum Associates.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, 44 (1), 190-204.

Newcomb, M. D., & Bentler, P. M. (1988). *Consequences of Adolescent Drug Use: Impact on the Lives of Young Adults.* Newberry Park: Sage.

Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment & Education in Higher Education, 29,* 3-19.

Onwuegbuzie, A. J. (1995). Statistics test anxiety and female students. *Psychology of Woman Quarterly, 19,* 413-418.

Petrides, K. V., Jackson,C. J., Furnham, A. & Levine, S. (2004). Development of a Short Form of the Eysenck Personality Profiler via Structural Equation Modeling, Kees van Montfort, Johan Oud and Albert Satorra (Eds.) *Recent Developments on Structural Equation Models*, Boston: Kluwer Academic Publishers, 221-239.

Pietsch, J., Walker, R., & Chapman, E. (2003). The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school, *Journal of Educational Psychology, 95,* 589–603.

Rensvold, R. B., & Cheung, G. W. (1999). Identification of influential cases in structural equation models using the jackknife method. *Organization Research Methods, 2*, 293-303.

Rindskopf, D. (1984). Structural equation models: Empirical identification, Heywood cases, and related problems, *Sociological Methods and Research,* 13, 109–119.

Rubin, D. B. (1987). *Multiple imputation for non-response in surveys.* New York: John Wiley.

Salzberger, T. (2006). Statistically equivalent, yet different in meaning: Some considerations on different measurement model representations. Retrieved October 2, 2008, from http://www2.wu-wien.ac.at/marketing/user/salzberger/research/aggloshop.pdf

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality & Social Psychology*, 46, 929-938.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analysis perspective. *Multivariate Behavioral Research, 33*(4), 545-571.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23-74.

Sluis, S., Dolan, C., Stoel, R. (2005). A note on testing perfect correlations in SEM. *Structural Equation Modeling, 12*(4), 551-557.

SPSS for Windows, Version 12.0.1 (2003). Australia: SPSS, Inc.

Sturgis, P. (2006). Surveys and Sampling. In *Research Methods in Psychology*, edited by G. Breakwell, S. Hammond and C. Fife-Schaw. London: Sage.

Thompson, B. (2003). *Score Reliability: Contemporary Thinking on Reliability Issues.* Thousand Oaks, CA: Sage.

Tomarken, A. J. & Waller, N.G., (2005). Structural Equation Modeling: Strengths, Limitations, and Misconceptions, *Annual review of clinical psychology*, 1, 31-65

Wood, M. (2005). Bootstrapped confidence intervals as an approach to statistical inference. *Organizational Research Methods, 8*(4), 454-470.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMMS data. *Practical Assessment, Research, & Evaluation, 12*(3), 1-26.

Yuan, K.H., & Chan, W. (2008). Structural equation modeling with near singular covariance matrices, *Computational Statistics and Data Analysis*, 52, 4842–4858.

Appendix A

Current Statistics Self-efficacy

Answer each question on a 1 (no confidence at all) to 5 (complete confidence scale)

1. Identify the scale of measurement for a variable.
2. Interpret the probability value (p-value) from a statistical procedure.
3. Identify if a distribution is skewed when given the values of three measures of central tendency.
4. Select the correct statistical procedure to be used to answer a research question.
5. Interpret the results of a statistical procedure in terms of the research question.
6. Identify the factors that influence power.
7. Explain what the value of the standard deviation means in terms of the variable being measured.
8. Distinguish between a Type I error and a Type II error in hypothesis testing.
9. Explain what the numeric value of the standard error is measuring.
10. Distinguish between the objectives of descriptive versus inferential statistical procedures.
11. Distinguish between the information given by the three measures of central tendency.
12. Distinguish between a population parameter and a sample statistic.
13. Identify when the mean, median and mode should be used as a measure of central tendency.
14. Explain the difference between a sampling distribution and a population

Appendix B

Statistics Anxiety Rating Scale

The following items refer to experience that may cause anxiety. Circle the number indicating the amount of anxiety you would experience with each of the situations. One (1) indicates no anxiety, and five (5) indicates very much anxiety.

1. Studying for an examination in a statistics course.
2. Interpreting the meaning of a table in a journal article
3. Going to ask my statistics teacher for individual help with material I am having difficulty understanding.
4. Doing the homework for a statistics course
5. Making an objective decision based on empirical data
6. Reading a journal article that includes some statistical analysis
7. Trying to decide which analysis is appropriate for your research project
8. Doing the final exam in a statistics course
9. Reading an advertisement for an automobile which includes figures on gas mileage, compliance with population regulations, etc.
10. Walking into the classroom to take a statistics test
11. Interpreting the meaning of a probability value once I have found it.
12. Arranging to have a body of data put into a computer
13. Finding that another student in the class got a different answer than you did to a statistical problem
14. Figuring out whether to reject or retain the null hypothesis
15. Waking up in the morning on a day of a statistics test
16. Asking one of your professors to help in understanding a printout
17. Trying to understand the odds in a lottery
18. Seeing a student poring over the computer printouts related to his/her research
19. Asking someone in the computer center for help in understanding a printout
20. Trying to understand the statistical analysis described in the abstract
21. Enrolling in a statistics course
22. Going over a final examination in statistics after it has been granted.
23. Asking a fellow student for help in understanding a printout.

Citation:

Larwin, Karen & Harvey, Milton (2012). A Demonstration of a Systematic Item-Reduction Approach Using Structural Equation Modeling. *Practical Assessment, Research & Evaluation*, 17(8). Available online: http://pareonline.net/getvn.asp?v=17&n=8

Authors:

Karen Larwin
Youngstown State University
khlarwin [at] ysu.edu

Milton Harvey
Kent State University
mharvey [at] kent.edu