The Beck Depression Inventory 2nd Edition (BDI-II): A Cross Sample Structural Analysis

Kamden K. Strunk

Auburn University

Forrest C. Lane

Sam Houston State University

NOTICE: This is the authors' version of a work that was accepted for publication in *Measurement and Evaluation in Counseling and Development*. Changes resulting from the publication process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may not have been made to this work since it was submitted for publication. This article may not exactly replicate the final published version. A definitive version was subsequently published in *Measurement and Evaluation in Counseling and Development*. doi: 10.1080/0781756.2017.1318636

Abstract

A common concern about the BDI-II among researchers in the area of depression has long been the single-factor scoring scheme. Methods exist for making cross sample comparisons of latent structure but tend to rely on estimation methods that can be imprecise and unnecessarily complex. This study presents a computationally more efficient method for reproducing a population matrix and uses this method to evaluate competing factor solutions of the BDI-II published in the literature. The results of this study suggest our alternative method can reproduce a population matrix and when used to test various factor solutions of the BDI-II, there is a lack of evidence to suggest any one superior structure. However, the source of measurement error within some models may help to inform the diagnostic utility of the BDI-II.

Keywords: Beck Depression Inventory-II; meta-analysis; confirmatory factor analysis

Testing Instrument Structure Across Multiple Studies: Illustration and Application using the
Beck Depression Inventory 2nd Edition

Advertised for the detection and assessment of depression, the Beck Depression Inventory, 2nd edition (BDI-II; Beck, Steer, & Brown, 1996) is widely used by clinicians in the screening and tracking of depression symptoms. It consists of 21 items that are summed to create a composite score of depression (Beck, Steer, & Brown, 1996). Examples of these items include questions regarding changes in sleep patterns, difficulty concentrating, sadness, self-dislike, crying, loss of energy, and suicidal thoughts. These items were designed to capture the depression as defined by the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (American Psychiatric Association, 2000).

A common concern about the BDI-II among researchers in the area of depression has long been the single-factor scoring scheme. This has led to a preponderance of exploratory and confirmatory factor analytic studies resulting in a variety of two, three, hierarchical and general factor solutions. Such controversy regarding BDI-II factors, as well as their structure, makes it a well-suited candidate for cross-sample structural analysis. The problem is that such comparisons tend to rely on estimation methods that may not fully reproduce a population matrix and can be unnecessarily complex. The purpose of this study was twofold: to present a computationally more efficient method for reproducing a population matrix; and to use this method to estimate population BDI-II data to evaluate competing factor solutions published in the literature.

BDI-II Factor Solutions

Among the two-factor solutions of the BDI-II, the cognitive/somatic split is most common in the literature. However, there are many variations on this general theme. Some include a somatic-affective factor alongside a cognitive factor (Arnau, Meagher, Norris, &

Bramson, 2001; Vanhuele, Desmet, Groenvynch, Rosseel, & Fontaine, 2008; Viljoen, Iverson, Griffiths, & Woodward, 2003). Others include a cognitive-affective factor paired with a somatic factor (Patterson, Morasco, Fuller, Indest, Loftis, & Hauser, 2011; Siegert, Walkey, & Turner-Stokes, 2009; Storch, Roberti, & Roth, 2004; Whisman, Perez, & Ramel, 2000). In one study, Wilson VanVoorhis and Blumentritt (2007) found a cognitive-somatic factor and an affective factor. Still others follow a simple cognitive/somatic split (Grothe, Dutton, Jones, Bodenlos, Ancona, & Brantley, 2005; Palmer & Binks, 2008; Quilty, Zhang, & Bagby, 2010; Thombs, Ziegelstein, Beck, & Pilote, 2008), though one labels this differently as cognitive and non-cognitive (Steer, Ball, Ranieri, & Beck, 1999). Brown, Kalpan, and Jason (2011) report a somatic-affective and role-emotional two-factor solution, deviating from the solutions reported by other researchers. Even among these studies, there is wide variation in the number of items per factor and which items are included on each factor. Thus, even if a two-factor solution is to be preferred, it remains unclear as to what two-factor structure may perform best.

Other researchers have found evidence in favor of three-factor solutions. These vary more widely in their structure and interpretation than do the two-factor solutions. There is some convergence around the idea of cognitive, affective, and somatic factors (Brouwer, Meijer, & Zevalkink, 2012; Johnson, Neal, Brems, & Fisher, 2006; Lindsay & Skene, 2007; Tully, Winefield, Baker, Turbull, & de Jonge, 2011; Vanhuele, et al., 2008). Yet, many other solutions may be found in the literature as well. For example, Byrne, et al. (2007) suggests negative attitude, performance difficulty, and somatic elements, though the items on the factors deviate only somewhat from other solutions interpreted as cognitive, affective, and somatic. Chilcot, Norton, Wellsted, Almond, Davenport, and Farrington (2011) suggest a cognitive, self-criticism, and anhedonia solution that differs substantially from other solutions. Lopez, Pierce, Gardner,

and Hanson (2012) found a three factor solution interpreted as negative rumination, somatic complaints, and mood that also differs from the other three factor solutions in general structure of the items.

Researchers have also suggested hierarchical and general factor solutions. Within the general factor solutions, the researchers add an additional factor for “depression” onto which all items individually load. In hierarchical solutions, the researchers add an additional factor for “depression” onto which all first-order factors load. Both Quilty, Zhang, and Bagby (2010) and Thombs, et al. (2008) found the best fit with a general depression factor onto which all items load in their two-factor solutions, as did Chilcot, et al. (2011) with a three-factor solution plus the general factor. Byrne, et al. (2007) has suggested a hierarchical solution on top of the three-factor model, while Grothe, et al. (2005) has suggested likewise on the two-factor model.

These studies raise real questions about the structure of the BDI-II. It is possible that structural differences may be a result of sample characteristics that may affect item loadings. For example, Bos, et al. (2009) found that the structure was variant in women between pregnancy and postpartum measurement. Carmody (2005) found that three factors fit best among college students, as did Johnson, Neal, Brems, and Fisher (2006) among injecting drug users. However, Cohen (2007) found evidence using multidimensional scaling that the structure was two-dimensional among the general population. A four-factor solution was reported among Puerto-Rican elderly participants (Rodriguez-Gomez, Davila-Martinez, & Collazo-Rodrigues, 2006). Ward (2006) suggested that three-factor models explain incrementally more variance, and that the practical interpretation of two-factor models may be misleading. Yet, Vanhuele, et al. (2008) found two and three factor models fit equally well among a mixed sample of clinical and non-

clinical participants. If the structure of the BDI-II fluctuates depending on the sample for a given study, how will it be structured in the population at large?

Population Matrix Estimation

Questions about the BDI-II structure in the population can be tested by using sample covariance or correlation matrices published in the literature. Methods exist for synthesizing these matrices in meta-analytic structural equation modeling (MA-SEM) applications. Confirmatory factor analysis is part of structural equation modeling, so it is natural that MA-SEM methods could be applied to questions of instrument structure across aggregated samples. Because MA-SEM techniques are designed to work in cases where not all variables will be present in all samples, they necessarily use estimation techniques to produce a combined covariance matrix.

One such approach when conducting MA-CFA is presented in Beretvas and Furlow (2009). This approach requires separate synthesis estimates for standard deviations and correlations, which are then converted to covariance matrices. This separate synthesis accounts for the fact that all information may not be present in all analyses. However, in the case of confirmatory factor analysis (CFA), this should not be necessary given all elements should be present in published studies testing the instrument's factor structure. This may also be unnecessary given that correlations and standard deviations can be converted into covariance matrices, and the covariance estimates synthesized.

The second issue concerning MA-SEM techniques that they rely on generalized least squares estimation and often involve very complex multi-step matrix algebra procedures for estimation, some of which require the download of specialized software (for example, software that creates EQS or LISREL code). A method that produces exact combined covariance matrices

and which is procedurally simpler would benefit the field by making it easier for researchers and practitioners interested in producing an assessment of structure across samples to do so. Also, by simplifying the analysis, such techniques may better encourage structural analyses and lead to a better understanding of psychological assessment instruments, in general. This understanding is particularly important in cases where instrument structure has been controversial (e.g., BDI-II), and where the assessment instrument has an impact on quality of care, diagnoses, and treatment.

We suggest published studies on factor structure containing inter-item correlation matrices can be used to estimate a population matrix. These correlations can be converted to variance ($s_x^2 = (s_x)^2$)/covariance ($s_{xy} = r_{xy}(s_x s_y)$) matrices that can then be converted into a Sum of Squares and Cross Product deviation (SSCP) matrix ($\sum(X - \bar{X})(Y - \bar{Y}) = s_{xy}(N - 1)$). The SSCP matrix contains the sum of squared deviation scores on the diagonal, and the sum of cross product deviation scores on the off-diagonal. Because SSCP values are additive, it is then possible to add together the values for all SSCP matrices, and divide by the total combined sample size minus one to obtain the combined variance/covariance matrix (Figure 1).

The procedure for creating the combined variance covariance matrix can easily be completed using computer software (e.g. Microsoft Excel, the SPSS matrix subprogram), or by hand. For example, in Microsoft Excel, it is possible to enter the standard deviations along with inter-item correlation coefficients in a worksheet, then use Excel formulae to calculate a variance/covariance and SSCP matrix. Each correlation coefficient is multiplied by the product of the standard deviations of each item ($r_{X1X2}(s_{X1}s_{X2})$) with variance calculated as the standard deviation squared ($(s_{X1})^2$). The variance/covariance matrix can then be converted to the SSCP matrix by multiplying each element by $n - 1$ ($s_{X1X2}(N - 1)$ and $s_{X1}^2(N - 1)$). A full diagram illustrating the necessary calculations is found in Figure 1. This procedure is repeated for each

sample (i.e., each study included in the meta-analysis). The combined SSCP matrix is created by summing all SSCP matrices from each sample ($\sum X_i'X_i$). The combined variance/covariance matrix is created by dividing the combined SSCP matrix values by the total sample size of all samples combined ($\frac{\sum X_i'X_i}{N_{total}-1}$). In the present study, Microsoft Excel was used to convert and combine the matrices.

Other software could also be used to accomplish the same goal. The correlation matrices can also be converted to variance/covariance matrices using the SPSS MCONVERT function. The SPSS COMPUTE function can then be used to convert to SSCP matrices, and to combine all matrices. Once a combined variance/covariance matrix is obtained, it can be analyzed in any structural modeling software package that accepts covariance matrices as input (e.g. Mplus, AMOS). This approach is computationally simpler and deviates from approaches suggested in Beretvas and Furlow (2009), Cheung and Chan (2009), and Furlow and Beretvas (2005) in that it uses neither generalized least squares estimation or multiple synthesis procedures.

The purpose of this paper is to present empirical evidence of this covariance matrix pooling approach to replicate a population matrix and illustrate use of this approach in real context. Specifically, various proposed models for the BDI-II were tested by pooling the variance-covariance matrix representing the population structure of the instrument. The following research questions guided this study: First, would the simulation method adequately replicate the population matrix in simulated data? Then, if the procedure is adequate, will any particular structure or set of structures stand out among those that have been proposed for the BDI-II? And finally, will those structures be dependent on clinical versus non-clinical populations?

Method

Data Simulation of Matrix Estimation Procedure

Because this method uses an alternative approach not well-tested in the literature, data simulation was used to evaluate how well such a method might reproduce the population covariance matrix. This simulation concentrated on establishing the adequacy of the proposed procedure for pooling SSCP matrices given the conversion between covariance and correlation is well established. A publicly available structure (Hoelzle, 2010) was used to generate an empirical population covariance matrix with one million cases. Population data was specified to be multivariate normal using the *MASS* package (Genz, Bretz, Miwa, Mi, Leisch, Scheipl, Bornkamp, & Hothorn, 2013) in R version 3.0.2 (R Core Team, 2013). All conditions were fully crossed and replicated 1,000 times.

Simulation Conditions. Previously published MA-CFA studies were used to guide the identification of conditions likely to influence the replication of population covariance matrices (Beretvas & Furlow, 2006; Cheung & Chan, 2009). The first variable contained two (2) levels of population variance. One level contained data using the specified variances from the original covariance matrix. The second level was a new population with an identical covariance structure but with item variances twice the original level. In addition, five (5) levels of single study samples size were considered in the analysis: 50, 100, 200, 500, and 1000. Lastly, four (4) levels representing the number of studies used in the pooling of covariance matrices were also considered within each population: 5, 10, 20, and 30 studies. Cases were randomly drawn with replacement and then used to estimate a covariance matrix among items for each simulated draw. This was done using a loop function in R given the specified study sample size for each level of population variance. In total, forty (40) conditions were simulated within the data.

Method for pooling covariance matrices. Pooled covariance matrices were obtained using the sums of squares and cross products matrices (SSCP) for each random sample. As each new sample was drawn, SSCP matrices were pooled into one covariance matrix (S) based on the total number of studies (NS) for that condition denoted as,

$$S = \frac{\sum(X_i'X_i)}{N-1},$$

where $X_i'X_i$ is the SSCP matrix and N is the total sample size within the condition.

Assessment of pooled matrices. The homogeneity of the combined (S) and specified population matrices were then tested using the one sample covariance test based on a log-likelihood ratio discussed in Mardia, Kent, and Bibby (1979). This test is expressed as,

$$-2\log\lambda = n(\text{tr}\{\Sigma_0^{-1}S\} - n(\log|\Sigma_0^{-1}S|) - np),$$

where n is the sample size, Σ_0 is the specified population covariance matrix, and p is number of variables in the matrix with an asymptotic chi-square distribution with $\frac{1}{2}p(p + 1)$ degrees of freedom. The homogeneity of S was tested using a criterion of $\alpha = .05$. Both the calculated test statistics and p value were saved for each simulated case. The R code used for the one sample covariance test is found in Appendix 1. All data were tested in a three-way (2x5x4) ANOVA. Both main effects and the cross-level interaction were examined, considering both p values and effect sizes in the interpretation of results (Henson, 2006; Vacha-Haase & Thompson, 2004; Wilkinson & APA Task Force on Statistical Inference, 1999).

BDI-II Data Sources

Data were then collected from published factor analyses of the BDI-II inter-item correlation matrices following evaluation of the matrix pooling approach. Authors of other factor analytic work with the BDI-II where the correlation matrix was not published were also contacted and asked for copies of the inter-item correlation matrix with standard deviations,

though relatively few provided these data due to availability, changes in institutional affiliation, lost data, or changes in email address making it impossible to contact them. Studies were gathered from databases including Academic Search Elite, Education Source, ERIC, HealthSource, MEDLINE, PsycArticles, Psychology and Behavioral Sciences Collection, PsycInfo, and SocIndex. Initially, 27 studies were identified that included a factor analysis of BDI-II. In total, 10 studies were included in the final data set based on those which included an inter-item correlation matrix or whose authors were able to produce or locate one. All ten studies included in the analysis and their sample characteristics are summarized in Table 2. The average sample size was 501.3, with a total combined sample size of 5,013. Seven of the samples come from clinical populations, while three come from non-clinical (student) populations. Although these studies include samples with varied characteristics, they are combined in this case in an attempt to approach the population as a whole, rather than any subset thereof. As a result, for this study, the “population” is thought of as all individuals who may be assessed with the BDI-II, both depressed and not depressed, of all age groups, and all ethnicities.

Analysis of Pooled Covariance Matrices

Individual correlation matrices from the samples above were combined into a pooled covariance matrix. Some precision may have been lost in the pooling procedure due to the fact that most published correlation matrices include only two decimal places. However, this loss was judged to be acceptable as it was at the hundredths place (i.e., .001 - .009) and was unlikely to meaningfully affect the interpretation of overall model fit. The combined overall variance/covariance matrix was used in all structural analyses. The adequacy of model fit was considered using model χ^2 , Tucker-Lewis index (TLI), comparative fit index (CFI), standardized root mean square residual (SRMR), and the root mean square error of

approximation (RMSEA). Fit indices were selected for inclusion based on their regular use in the confirmatory factor analytic literature, and because of their differential methods of testing fit (Kline, 2011; Schreiber, Nora, Stage, Barlow, & King, 2006).

Results

Matrix Estimation Procedure

Simulated data were first examined to evaluate the ability of the method to replicate a population matrix. Mean and standard deviations for simulated data conditions are reported in Table 1. The mean one-sample covariance test statistic for all simulated data was 15.407 ($SD = 5.67$) with an associated p value of .480 ($SD = .290$). This suggested no difference between the pooled estimate of the population matrix and the actual population matrix used in the testing of this approach (Hoelzle, 2010). Box's M values decreased as the sample size improved. This result seemed appropriate given that covariance matrices were being estimated from sample sizes that should better approximate the true population parameter. In contrast, Box's M values generally increased as more studies were used to create a pooled covariance matrix. Pooled covariance matrices under these conditions represent a larger combined sample size and we would expect statistical tests to be increasingly sensitive to such change.

Because sample means varied across conditions, this data was tested the data using three-way ANOVA. Omnibus ANOVA results indicated no statistically significant interaction effects but the main effects for population variance ($F_{1, 39960} = 6.247, p = .012, \eta^2 = .001$), number of pooled covariance matrices ($F_{3, 39960} = 23.476, p < .001, \eta^2 = .001$), and single study sample size ($F_{4, 39960} = 40.590, p < .001, \eta^2 = .004$) were all statistically significant ($\alpha = .05$). These differences were notably small ($\eta^2 < 1\%$) but were examined using planned contrasts to better inform specific group differences. Contrasts were specified such that smallest condition served

as a basis for comparison. Box's M was statically larger when either 15 ($b = .309, p < .001$) or 30 matrices ($b = .624, p < .001$) were pooled together to reproduce a population covariance matrix. Each increasing level of single study sample size also resulted in statistical decreases to Box's M ([$n=100$] $b = -.616, p < .001$; [$n = 200$] $b = -.883, p < .001$; [$n=500$] $b = -.935, p < .001$; [$n=1,000$] $b = -.964, p < .001$). However, none of the conditions simulated resulted in a p values less than $\alpha = .05$ for any of the one-sample tests of covariance matrices. All conditions statistically approximated of the population covariance matrix. This seemed to suggest this procedurally simpler method for pooling matrices was appropriate in the statistical reproduction of the population covariance matrix.

BDI-II Factor Solutions from Pooled Sample Data

Given the results of the matrix estimation procedure discussed above, the ten BDI-II studies in which an inter-item correlation matrix was included or whose authors were able to produce or locate one were used to estimate a population variance/covariance matrix. Confirmatory factor analyses were conducted on this covariance matrix in Mplus Version 6.11. A total of 21 models were tested from the literature, including nine two-factor models, six three-factor models, and six models with hierarchical or general factors (see Table 3 for a summary of all models tested). The results of the initial confirmatory factor analyses can be found in Table 4. All models fit relatively well in terms of the Confirmatory Fit Index (CFI), Tucker Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA). With one exception, all CFI values were above .90, as were all TLI values. In addition, all RMSEA values were between .04 and .06. However, in all cases the chi-square to degrees of freedom ratio was unacceptable. One potential issue with the interpretation of this result is that the chi-square statistic becomes an increasingly poor indicator of fit as the sample size increases, making other fit indices more

relevant for interpretation (Byrne, 2012; Kline, 2011). Because of this issue, modification and refitting of models is also a less useful proposition. This is particularly true given that the nature of the present inquiry is to determine the structure of the instrument given that many different latent structures have been proposed, so the theoretical rationale for modifications in the setting of this study is lacking, particularly given that such modifications as error covariances often do not generalize and generalization is the goal of the present study.

To address the question of whether structure may be based on the clinical nature of the population, two separate combined group covariance matrices were created: one for the student, or “normal” samples, and another for the clinical samples. Using AMOS Version 20.0, a multigroup confirmatory factor analysis was conducted. The model selected for this preliminary analysis was that of Thombs, et al. (2008), primarily because Beck was a coauthor on that study. The basic question addressed in this analysis was whether there would be differences in the structure of the instrument for clinical versus non-clinical samples.

Using a chi-square difference test, invariance was tested from the unconstrained model to one with measurement weights constrained, then one with structural covariances constrained, and finally a model with measurement residuals constrained, in nested models. The difference between the unconstrained model to the one with measurement weights constrained was significant ($\Delta\chi^2_{15} = 95.45, p < .001, \Delta CFI = .004, \Delta RMSEA < .001$), as was the difference between the model with measurement weights constrained and that between the model with measurement weights constrained and that with structural covariances constrained ($\Delta\chi^2_5 = 167.78, p < .001, \Delta CFI = .008, \Delta RMSEA = .002$), and the difference between the model with structural covariances constrained and the model with measurement residuals constrained ($\Delta\chi^2_{25}$

= 1078.48, $p < .001$, $\Delta CFI = .049$, $\Delta RMSEA = .011$). For a full listing of fit indices for each of the nested models, refer to Table 5.

However, the chi-square difference test has the potential to be misleading in large samples, so the difference in CFI was also assessed, where a difference of .02 has been suggested as a meaningful difference (Byrne, 2008; Cheung & Rensvold, 2002). Based on the difference in CFI, it appears that the two groups are equivalent in terms of measurement weights and structural covariances, but not at the level of measurement residuals. This suggests that the groups load similarly onto the factors, but that there are varying levels of error variance in the two groups. If indeed there are differing levels of error variance, the two groups would not be truly parallel in the classic sense, and these differences in error variance may explain the diversity of factor structure supported by researchers working in different settings.

Discussion

In the simulation study, the method adequately reproduced the simulated population covariance matrix. Then, in the application study, the method allowed for empirical evaluation of models in a larger, more diverse sample that was intended to more closely approximate the general population than any of the individual study's specialized samples could. This additionally allowed for analysis of the differences in structure between types of samples. The usefulness of the method was demonstrated through its application to the BDI-II. As a result, discussion of findings on the BDI-II are intended to further highlight the ways in which this method can benefit the study of instruments with controversial structure.

Method for Pooling Covariance Matrices across Samples

The method illustrated in this study for pooling covariance matrices has clear advantages – a side-by-side comparison of many published factor models can be created among shared data.

This puts the models on equal ground, and gives a large, varied data set on which to test the fit of those models. In addition, the method does not involve the same complex estimation processes as existing methods for combining variance/covariance matrix. Rather, the entire procedure takes place within the matrix algebra framework used to derive the original SSCP and covariance estimates. Simulation results in the present study demonstrate the adequacy of this method for reproducing the population covariance matrix across a variety of conditions. The results from this study lend confidence to the assertion that by combining matrices through the proposed method, it is possible to test alternate structural models using confirmatory factor analysis in a much larger sample that will more closely approximate the population matrix.

A potential limitation to the approach illustrated in this study is the use of disparate samples in creating the combined covariance matrix. Each has very different sample characteristics, which has one of two possible outcomes: First, it may result in a closer approximation of the population matrix. However, it may also result in a highly variable matrix that includes more than one population. That is, these disparate samples may actually represent multiple populations. Will college students, substance abuse patients, outpatient medical patients, psychiatric patients, pain patients, people of various ages and ethnicities, etc., all be considered part of the U.S. population, or will some subsets of these groups comprise discrete populations to be assessed separately? Future research may focus on the suitability of covariance matrices for combination, and the implications of combining or not combining matrices on how the population is theoretically constructed.

This method for testing instrument structure across multiple samples may be useful in any case where there are multiple factor analytic studies with conflicting results. This is particularly important to consider when the instrument in question has, as the BDI-II, practical

implications for health, education, policy, or other areas with implications for the people who will be measured using the instrument. In such cases, understanding the structure of the instrument takes on an added importance and value, and having an additional tool for assessing structure in the population may prove extremely useful.

BDI-II

The BDI-II seemed to fit relatively well in several models, with the exception of the chi-square to degrees of freedom ratio. Although recommended cutoffs for this ratio are much lower than values obtained in these models, there are a number of possible explanations for these high values. A number of factors may influence chi-square values in situations of large sample size and fairly reliable measures. For example, Hammervold & Olsson (2012) found that even slightly misspecified models were highly likely to be rejected in the chi-square test, even with extremely large sample sizes. Given the controversy surrounding the BDI-II and the idea that its structure may be sample-dependent, it is likely that each model is slightly misspecified in a sense. On the other hand, reliable models also tend to produce larger chi-square values in very large samples, according to simulation studies by Miles and Shevlin (2007). In other words, it may be that the large chi-square to degrees of freedom ratios in this case are misleading.

In the case that one would choose to view the large chi-square to degrees of freedom ratios as artifacts of the large sample and the measurement properties of the BDI-II, the picture becomes even more troubling. It would appear that just about any published model (with the exception of the Wilson VanVoorhis, et al. [2007] model) fits relatively well. Obviously some models fit better than others, but there is not adequate evidence to suggest that there is one true and superior model at the population level for the BDI-II from an empirical standpoint, at least in terms of model fit.

However, in general, the hierarchical and general factor models did not seem to offer much in terms of fit. One notable exception is the Thombs, et al. (2008) model. This model had very good fit, and one of the lowest chi-square to degrees of freedom ratios overall. Moreover, it does appear that the current research supports the assertions of Byrne, et al. (2007), Ward (2006), and others, who have argued strongly for the superiority of three-factor solutions. It is worth noting that the advantage in fit for three-factor models was significant, but modest, with many differences barely exceeding the threshold for consideration of .01 difference in CFI.

Additionally, because the difference in fit is negligible in many cases, researchers may consider straying away from structures that deviate heavily from the BDI-II's original scoring or administration. For example, Chilcot, et al. (2011) and Patterson, et al. (2011) each use less than half of the original items. Given that the BDI-II is a commercial instrument unlikely to change its item content to match factor analytic work (particularly if that work is not in the majority) it may be more useful to explore structures that are more practically relevant by using the entire (or close to the entire) instrument, and focusing on structures that provide information relevant to clinicians. Particularly with instruments like the BDI-II that are in constant use for screening and patient care, it may be meaningful to focus on instrument structures that are focused on such issues, especially given the very small differences in fit for most models. For example, though Patterson, et al. (2011) reduced the items in the model to get better fit, the model has only a .02 advantage in CFI over the other two-factor models, which may lead one to question if the advantage in fit is worth the tradeoff in practicality.

Of additional interest was the invariance in structure but variance in error and latent variable covariance between clinical and non-clinical samples. This is suggestive of why the chi-square was so large in many of the models – wider degrees of error variance and latent variable

covariance would result in a poorer fit than in a more homogeneous population. However, from that same analysis, it is interesting to note the differences were at the level of error variance and latent variable covariance, suggesting the actual factor loadings were similar. Of course, this is not a suitable argument for measurement invariance, but suggests the level at which measurement variance occurs. It is worth noting, too, the problematic nature of different structure in clinical and non-clinical samples. The BDI-II is intended, according to the publisher, for the purposes of detecting and assessing depression (Beck, Steer, & Brown, 1996). If one develops a measure that only adequately assesses depression in those known to be depressed, and a different scoring system is needed for those who are not depressed, the instrument would have no diagnostic utility. The diagnosis would have to be known in order to score the instrument, which is subsequently used for diagnostic purposes, creating a kind of nonsensical system. Even though the BDI-II is often used as a screening tool where the total score is assessed (thus ignoring any factor structure that may exist), the idea that it is a very different scale when given to depressed versus euthymic individuals is deeply problematic. The problem is even deeper when the instrument is used for research purposes and understanding the functions of depression, depression treatments, and how depressive symptoms (and sets of depression symptoms like cognitive, somatic, and affective) operate in different samples. Instead, a model is needed that functions reasonably well for both populations and can thus be used for the purposes of determining which class an individual, in fact, belongs to. It is reasonable to ask this question for any instrument, as to whether its structure across samples matches its stated purpose and uses, and the present method may be useful in this regard.

The suitability of different models cannot be finally resolved wholly through structural analyses such as confirmatory factor analysis, however. That is, while structural analyses can

establish the model that fits best to observed data, structural analysis cannot determine how well a given structure (and the resultant scores) are able to discriminate between diagnoses, or predict behavior. The structural analyses alone cannot establish construct validity. Validity is a function of the interpretations of the scores, which require research beyond structural work alone. Empirical studies of construct validity for interpretations arising from particular models, especially for predictive validity, will be an important next step in evaluating alternative structures.

Of course, it is possible that none of the tested models actually represent the ‘true’ latent factor structure, thus creating a muddled picture with no clearly best-fitting model emerging. Given that all models posited in the factor-analytic literature were tested, this would represent a larger problem with the theorizing of depression and specifically the BDI-II wherein either the measure itself has been ill-conceived or the theorizing of its structure has been widely off-base. In this case, a new theory of the measure and how it fits with contemporary understandings of depression would be needed. This conclusion seems extreme, however, in light of the relatively good fit indices obtained in the present study. It may instead be the case that some slightly improved model exists that has not been posited that would better fit the data.

The present analysis of the structure for the BDI-II is limited, however. Although available studies have been synthesized in the present study, the nature of the available data is a limitation. Because 70% of the sampled studies were with clinical samples, and the remaining 30% were with non-clinical student samples, the combined covariance matrix does not fully represent the population of individuals who may be assessed with the BDI-II. There are many non-clinical populations who may be assessed with the BDI-II for a number of practical purposes such as screening, evaluation, or as part of a larger battery. These portions of the population are

unrepresented in the combined covariance matrix. Although combining covariance matrices has the potential to create a matrix that represents or approaches the population, this is entirely dependent on the qualitative nature of the available data sources. The results of this study may not generalize outside of the samples represented in the data sources.

Conclusion

The proposed method for pooling covariance matrices in meta-analytic confirmatory factor analysis adequately reproduced the population covariance matrix under a number of conditions in the simulation study. Although the degree of error increased under some conditions, the difference between the reproduced matrix and the population matrix was never statistically significant, suggesting the method is robust to differences in number of samples, sample sizes, and variance levels. Moreover, the method was illustrated using the BDI-II, leading to the conclusion that many models fit reasonably well in the pooled matrix, though some stood out as fitting somewhat better. In cases like the BDI-II, a method for meta-analytic confirmatory factor analysis can either bring clarity as to which factor structure should be preferred, or highlight the potential adequacy of several factor structures, thus calling for more methods of assessing the practical utility of those alternatives.

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Arnau, R. C., Meagher, M. W., Norris, M. P., & Bramson, R. (2001). Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. *Health Psychology, 20*(2), 112-119.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
- Beretvas, S. N., & Furlow, C. F. (2009). Evaluation of an approximate method for synthesizing covariance matrices for use in meta-analytic SEM. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(2), 153-185.
- Bos, S. C., Pereira, A. T., Marques, M., Maia, B., Soares, M., J., Valente, J., Gomes, A., Macedo, A., & Azevedo, M. H. (2009). The BDI-II factor structure in pregnancy and postpartum: Two or three factors? *European Psychiatry, 24*, 334-340.
- Brouwer, D., Meijer, R. R., & Zevalkink, J. (2012). On the factor structure of the Beck Depression Inventory-II: G is the key. *Psychological Assessment*. Advance online publication.
- Brown, M., Kaplan, C., & Jason, L. (2011). Factor analysis of the Beck Depression Inventory-II with patients with chronic fatigue syndrome. *Journal of Health Psychology, 17*(6), 799-808.
- Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument. A walking through the process. *Psicothema, 20*(4), 872-882.

- Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Routledge.
- Byrne, B. M., Steward, S. M., Kennard, B. D., & Lee, P. W. H. (2007). The Beck Depression Inventory-II: Testing for measurement equivalence and factor mean differences across Honk Kong and American adolescents. *International Journal of Testing*, 7(3), 293-309.
- Carmody, D. P. (2005). Psychometric characteristics of the Beck Depression Inventory-II with college students of diverse ethnicity. *International Journal of Psychiatry in Clinical Practice*, 9(1), 22-28.
- Cheung, M. W. L., & Chan, W. (2009). A two-stage approach to synthesizing covariance matrices in meta-analytic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(1), 28-53.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 13, 287-321.
- Chilcot, J., Norton, S., Wellsted, D., Almond, M., Davenport, A., & Farrington, K. (2011). A confirmatory factor analysis of the Beck Depression Inventory-II in end-stage renal disease patients. *Journal of Psychosomatic Research*, 71, 148-153.
- Cohen, A. (2007). The underlying structure of the Beck Depression Inventory II: A multidimensional scaling approach. *Journal of Research in Personality*, 42, 779-786.
- Furlow, C. F., & Beretvas, S. N. (2005). Meta-analytic methods of pooling correlation matrices for structural equation modeling under different patterns of missing data. *Psychological Methods*, 10(2), 227-254.

- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2014). *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9-9997. URL <http://CRAN.R-project.org/package=mvtnorm>
- Hammervold, R., & Olsson, U. H. (2012). Testing structural equation models: The impact of error variances in the data generating process. *Quality and Quantity*, 46, 1547-1570.
- Henson, R.K. (2006). Effect-size measures and meta-analytic thinking in counseling psychology research. *The Counseling Psychologist*, 35(5), 601-629.
- Hoelzle, B. (2010). *Monte Carlo simulation* [R code available from EDU 7312 - Intermediate Quantitative Statistics: Multiple Regression]. School of Education and Human Development, Southern Methodist University. Retrieved from http://faculty.smu.edu/kyler/courses/7312/Hoelzle_Monte_Carlo.txt.
- Grothe, K. B., Dutton, G. R., Jones, G. N., Bodenlos, J., Ancona, M., & Brantley, P. J. (2005). Validation of the Beck Depression Inventory – II in a low-income African American sample of medical outpatients. *Psychological Assessment*, 17(1), 110-114.
- Johnson, M. E., Neal, D. B., Brems, C., & Fisher, D. G. (2006). Depression as measured by the Beck Depression Inventory-II among injecting drug users. *Assessment*, 13(2), 168-177.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford.
- Lindsay, W. R., & Skene, D. D. (2007). The Beck Depression Inventory II and the Beck Anxiety Inventory in people with intellectual disabilities: Factor analyses and group data. *Journal of Applied Research in Intellectual Disabilities*, 20, 401-408.

- Lopez, M. N., Pierce, R. S., Gardner, R. D., & Hanson, R. W. (2012, September 24). Standardized Beck Depression Inventory-II Scores for Male Veterans Coping with Chronic Pain. *Psychological Services*. Advance online publication.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate Analysis*. New York, NY: Academic Press.
- Miles, J., & Shevlin, M. (2007). A time and a place for incremental fit indices. *Personality and Individual Differences*, 42(5), 869-874.
- Osman, A., Barrios, F. X., Gutierrez, P. M., Williams, J. E., & Bailey, J. (2008). Psychometric properties of the Beck Depression Inventory-II in nonclinical adolescent samples. *Journal of Clinical Psychology*, 64(1), 83-102.
- Palmer, E. J., & Binks, C. (2008). Psychometric properties of the Beck Depression Inventory-II with incarcerated male offenders aged 18-21 years. *Criminal Behaviour and Mental Health*, 18, 232-242.
- Patterson, A. L., Morasco, B. J., Fuller, B. E., Indest, D. W., Loftis, J. M., & Hauser, P. (2011). Screening for depression in patients with hepatitis C using the Beck Depression Inventory-II: Do somatic symptoms comprise validity? *General Hospital Psychiatry*, 33, 354-362.
- R Development Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rodriguez-Gomez, J. R., Davila-Martinez, M. G., & Collazo-Rodriguez, L. C. (2006). Factor structure of the Beck Depression Inventory-Second Edition (BDI-II with Puerto Rican Elderly. *Puerto Rican Health Sciences Journal*, 25(2), 127-132.

- Quilty, L. C., Zhang, K. A., & Bagby, R. M. (2010). The latent symptom structure of the Beck Depression Inventory-II in outpatients with major depression. *Psychological Assessment*, 22(3), 603-608.
- Schreiber, J. B., Stage, F. K., King, J., Nora, A., Barlow, E. A., King, J. (2006) Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, 99(6), 323-337.
- Siebert, R. J., Walkey, F. H., & Turner-Stokes, L. (2009). An examination of the factor structure of the Beck Depression Inventory-II in a neurorehabilitation inpatient sample. *Journal of the International Neuropsychological Society*, 15, 142-147.
- Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology*, 55(1), 117-128.
- Storch, E. A., Roberti, J. W., & Roth, D. A. (2004). Factor structure, concurrent validity, and internal consistency of the Beck Depression Inventory-Second Edition in a sample of college students. *Depression and Anxiety*, 19, 187-189.
- Thombs, B. D., Ziegelstein, R. C., Beck, C. A., & Pilote, L. (2008). A general factor model for the Beck Depression Inventory-II: Validation in a sample of patients hospitalized with acute myocardial infarction. *Journal of Psychosomatic Research*, 65, 115-121.
- Tully, P. J., Winefield, H. R., Baker, R. A., Turnbull, D. A., & de Jonge, P. (2011). Confirmatory factor analysis of the Beck Depression Inventory-II and the association with cardiac morbidity and mortality after coronary revascularization. *Journal of Health Psychology*, 16(4), 584-595.

- Vacha-Haase, T. & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51*, 473-481.
- Ward, L. C. (2006). Comparison of factor structure models for the Beck Depression Inventory-II. *Psychological Assessment, 18*(1), 81-88.
- Whisman, M. A., Perez, J. E., & Ramel, W. (2000). Factor structure of the Beck Depression Inventory – Second Edition (BDI-II) in a student sample. *Journal of Clinical Psychology, 58*(4), 545-551.
- Wilkinson, L., & American Psychological Association Task Force of Statistical Inference. (1999). *Statistical methods in psychology journals: Guidelines and explanation. American Psychologist, 54*, 594-604. [reprint available through the APA Home Page: <http://www.apa.org/journals/amp/amp/548594.html>].
- Wilson VanVoorhis, C. R., & Blumentritt, T. L. (2007). Psychometric properties of the Beck Depression Inventory-II in a clinically-identified sample of Mexican American adolescents. *Journal of Child and Family Studies, 16*, 789-798.
- Vanheule, S., Desmet, M., Groenvynck, H., Rosseel, Y., & Fontaine, J. (2008). The factor structure of the Beck Depression Inventory-II. *Assessment, 15*(2), 177-187.
- Viljoen, J. L., Iverson, G. L., Griffiths, S., & Woodward, T. S. (2003). Factor structure of the Beck Depression Inventory-II in a medical outpatient sample. *Journal of Clinical Psychology in Medical Settings, 10*(4), 289-291.

Table 1

Means and Standard Deviations of Test Statistics and p values for Study Conditions

δ^2	NS	n	Box's M		p values	
			M	SD	M	SD
1	5	50	15.742	5.796	.460	.294
		100	15.351	5.780	.484	.288
		200	14.697	5.987	.518	.296
		500	15.191	5.821	.489	.282
		1000	15.052	5.592	.500	.289
		M (NS=5)	15.207	5.795	.490	.290
	10	50	15.685	5.864	.469	.297
		100	15.363	5.844	.487	.294
		200	15.109	6.020	.494	.297
		500	15.280	5.441	.490	.288
		1000	14.923	5.469	.504	.287
		M (NS=10)	15.272	5.728	.489	.293
	15	50	16.172	5.596	.443	.288
		100	15.725	5.750	.466	.285
		200	15.314	5.438	.485	.286
		500	15.238	5.801	.492	.296
		1000	14.937	5.717	.503	.295
		M (NS=15)	15.477	5.660	.478	.290
	30	50	16.707	5.806	.412	.295
		100	16.086	5.720	.450	.298
		200	15.871	5.399	.457	.291
		500	15.429	5.347	.481	.285
		1000	15.669	5.721	.466	.287
		M (NS=30)	15.953	5.599	.453	.291
2	5	50	15.579	5.714	.469	.291
		100	15.063	5.955	.495	.296
		200	14.983	5.896	.499	.289
		500	14.795	5.804	.512	.285
		1000	15.000	5.585	.495	.290
		M (NS=5)	15.084	5.791	.494	.290
	10	50	15.936	5.302	.456	.282
		100	15.114	5.792	.493	.294
		200	15.230	5.539	.490	.283
		500	14.954	5.353	.504	.285
		1000	14.972	5.676	.505	.292
		M (NS=10)	15.241	5.532	.490	.287
	15	50	16.115	5.582	.448	.288
		100	15.563	5.678	.473	.293
		200	14.991	5.672	.502	.297
		500	15.123	5.653	.489	.297
		1000	15.366	5.302	.480	.284
		M (NS=15)	15.432	5.577	.478	.292
	30	50	16.736	5.578	.414	.295
		100	15.479	5.286	.473	.283
		200	15.412	5.674	.477	.290
		500	15.263	5.622	.486	.289
		1000	15.045	5.320	.495	.284
		M (NS=30)	15.587	5.496	.469	.288

Table 2

Characteristics of Studies Included in Combined Covariance Matrix

Study	N	Gender		Age M(SD)	African- American	Asian	Ethnicity Hispanic /Latino	Native American	White	Population	C
		Men	Women								
Brouwer, Meijer, & Zevalkink (2012)	1530	61.5	38.5	35.1(10.1)	NR	NR	NR	NR	NR	Outpatient psychiatric clients	C
Brown, Kaplan, & Jason (2011)	111	17.1%	82.9%	44.1(11.6)	3.6%	3.6%	4.5%	NR	88.3%	Chronic Fatigue Syndrome patients	C
Carmody (2005)	502	45.8%	54.2%	20.18(?)	10%	7%	22%	2%	59%	Undergraduate students	N
Grothe, Dutton, Jones, Bodenlos, Ancona, & Brantley (2005)	220	48%	52%	49.26(10.91)	100%	0%	0%	0%	0%	African-American medical outpatients	C
Johnson, Neal, Brems, & Fisher, (2006)	598	75.8%	24.2%	38.2(7.8)	18.6%	1%	3.3%	20.1%	55.6%	Injecting drug users	C
Lopez, Pierce, Gardner, & Hanson (2012)	345	100%	0%	53.26(11.39)	11%	NR	9%	NR	49.3%	Chronic pain patients in the VA system	C
Osman, Barrios, Gutierrez, Williams & Bailey (2008)	414	50.7%	49.3%	15.77(1.03)	8.7%	NR	4.6%	NR	81.4%	High school adolescents	N
Steer, Ball, Ranieri, & Beck (1999)	210	50%	50%	41.29(15.25)	5%	2%	<1%	NR	93%	Outpatients diagnosed with depressive disorders	C
Thombs, Ziegelstein, Beck, & Pilote (2008)	477	82.6%	7.4%	60.1(9.2)	NR	NR	NR	NR	95.2%	Patients with acute myocardial infarction	C
Whisman, Perez, & Ramel (2000)	606	42.2%	57.8%	18.64(1.12)	6.3%	20.6%	6.3%	NR	60.2%	Undergraduate students	N

Note. C = clinical sample and N = nonclinical sample for the purposes of the multigroup analysis.

Table 3
Models Tested

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.
Arnau, et al.	SA	C	C	SA	C	C	C	-	C	C	SA	SA	SA	C	SA	SA	SA	SA	SA	SA	SA
Palmer, et al.	C	C	C	C	C	C	C	C	C	C	S	S	C	C	S	S	S	S	S	S	-
Patterson, et al.	-	CA	CA	-	CA	-	CA	CA	-	-	-	-	-	CA	S	-	-	S	S	S	-
Siebert, et al.	CA	CA	CA	S	CA	CA	CA	CA	CA	CA	CA	S	S	CA	S	S	-	S	S	S	-
Storch, et al.	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	S	S	CA	S	S	S	CA
Viljoen, et al.	SA	C	C	SA	C	C	C	C	C	C	SA	SA	-	C	SA	SA	SA	-	SA	SA	SA
Whisman, et al.	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	S	S	CA	S	S	-	CA
Wilson VanVoorhis, et al.	CS	CS	CS	CS	CS	CS	A	CS	A	CS	A	CS	A	-	A	A	A	A	A	A	A
Vanhuele, et al.	C	C	C	SA	-	C	C	-	C	SA	SA	-	-	C	SA	SA	-	SA	SA	-	SA
Brouwer, et al.	A	A	C	A	C	C	C	C	A	S	S	A	C	C	S	S	S	S	S	S	S
Byrne, et al.	NA	NA	NA	PD	NA	NA	NA	NA	NA	NA	PD	PD	PD	NA	S	S	PD	S	PD	S	-
Johnson, et al.	C	C	C	A	C	C	C	C	C	A	S	A	A	C	S	S	S	S	S	S	S
Lindsay, et al.	C	C	C	CA	C	C	C	C	C	C	C	C	CA	CA	C	CA	S	C	S	CA	S
Lopez, et al.	M	M	NR	S	NR	NR	NR	NR	-	NR	M	S	-	NR	S	-	M	-	S	S	S
Vanhuele, et al.	-	C	C	A	-	C	-	C	C	A	-	A	-	C	-	A	A	A	A	A	A
Byrne, et al. (H)	NA	NA	NA	PD	NA	NA	NA	NA	NA	NA	PD	PD	PD	NA	S	S	PD	S	PD	S	-
Chilcot, et al. (G)	-	C	C	An	C	C	SC	SC	C	-	-	An	-	C	-	-	-	-	-	-	-
Grothe, et al. (H)	C	C	C	S	C	C	C	C	C	S	S	S	S	C	S	S	S	S	S	S	S
Quilty, et al. (G)	-	S	S	-	S	S	S	S	S	-	-	-	-	S	C	C	-	C	C	C	-
Steer, et al. (H)	NC	C	C	NC	C	C	C	C	C	NC	NC	NC	NC	C	NC	NC	NC	NC	NC	NC	NC
Thombs, et al. (G)	C	C	C	-	C	C	C	C	C	-	-	-	-	C	S	S	S	S	S	S	S

Note. C = Cognitive, NC = Non-Cognitive, S = Somatic, A = Affective, SA = Somatic-Affective, CA = Cognitive-Affective, NA = Negative Attitude, PD = Performance Difficulty, NR = Negative Rumination, M = Mood, SC = Self-Criticism, An = Anhedonia. H = Hierarchical model with all factors loading onto a second level factor. G = General depression factor with all items, including items not loading on another factor, loading on a general depression factor.

Table 4

Summary of Model Fit Indices

	Model	χ^2	df	χ^2/df	CFI	TLI	RMSEA	SRMR
Two-Factors	Arnau, et al.	2674.71	169	15.83	.92	.91	.05	.04
	Palmer, et al.	3048.51	169	18.04	.91	.90	.06	.04
	Patterson, et al.	540.84	34	15.91	.97	.96	.06	.04
	Siebert, et al.	2643.12	151	17.50	.92	.91	.06	.04
	Storch, et al.	3500.65	188	18.62	.90	.89	.06	.04
	Viljoen, et al.	2547.39	151	16.87	.92	.91	.06	.04
	Whisman, et al.	3056.69	169	18.09	.91	.90	.06	.04
	Wilson VanVoorhis, et al.	3833.61	169	22.68	.88	.87	.07	.05
	Vanhuele, et al.	1216.50	103	11.81	.95	.94	.05	.03
Three-Factors	Brouwer, et al.	2664.48	186	14.33	.93	.92	.05	.04
	Byrne, et al.	2168.77	167	12.99	.94	.93	.05	.04
	Johnson, et al.	2475.15	186	13.31	.93	.93	.05	.04
	Lindsay, et al.	2238.11	132	16.96	.93	.91	.06	.04
	Lopez, et al.	1911.06	116	16.48	.94	.92	.06	.04
	Vanhuele, et al.	770.56	87	8.86	.97	.96	.04	.03
H & G Factors	Byrne, et al. (H)	2168.77	167	12.99	.94	.93	.05	.04
	Chilcot, et al. (G)	2441.18	179	13.64	.93	.92	.05	.04
	Grothe, et al. (H)	2859.01	188	15.21	.92	.91	.05	.04
	Quilty, et al. (G)	1760.25	176	10.00	.95	.95	.04	.03
	Steer, et al. (H)	1855.40	188	15.19	.92	.91	.05	.04
	Thombs, et al. (G)	1605.88	173	9.28	.96	.95	.04	.03

Note. H = Hierarchical factor, G = General factor.

Table 5

Fit Indices for Nested Models by Increasing Constraints

	χ^2	df	$\Delta\chi^2$	CFI	ΔCFI	RMSEA	
Unconstrained model	1164.130	170		.954		.034	
Measurement weights constrained	1259.582	185	95.452	.950	.004	.034	.000
Structural covariances constrained	1427.361	190	167.779	.942	.008	.036	.002
Measurement residuals constrained	2505.842	205	1078.481	.893	.049	.047	.011

Given the correlation matrix:

Correlation Matrix			
	X1	X2	X3
X1	-		
X2	r_{X1X2}	-	
X3	r_{X1X3}	r_{X2X3}	-

The covariance matrix will be calculated as:

Covariance Matrix			
	X1	X2	X3
X1	$(s_{X1})^2$		
X2	$r_{X1X2}(s_{X1}s_{X2})$	$(s_{X2})^2$	
X3	$r_{X1X3}(s_{X1}s_{X3})$	$r_{X2X3}(s_{X2}s_{X3})$	$(s_{X3})^2$

And the SSCP Matrix will be calculated as:

Sum of Squares and Cross Product (SSCP) Matrix = $X_i'X_i$			
	X1	X2	X3
X1	$s^2_{X1}(N - 1)$		
X2	$s_{X1X2}(N - 1)$	$s^2_{X2}(N - 1)$	
X3	$s_{X1X3}(N - 1)$	$s_{X2X3}(N - 1)$	$s^2_{X3}(N - 1)$

The combined SSCP matrix will be calculated as: $\sum X_i'X_i$

And the combined covariance matrix as: $\frac{\sum X_i'X_i}{N_{total}-1}$

Figure 1.

Process of Combining the Matrices

Appendix 1

R Syntax for the One Sample Covariance Test

```
cov.equal=function(x, Sigma,a=.05) {  
  ##x is the set of data  
  ##Sigma is the assumed covariance matrix  
  ##a is the level of significance set by default to .05  
  p= 5 ##dimensionality of the data  
  n1= n*i ##total sample size across all randomly drawn datasets  
  S=comb.cov  
  ##the next two lines construct the test statistic  
  mesa=solve(massive.cov)%*%S  
  test=n1*sum(diag(mesa))-n1*log(det(mesa))-n1*p  
  df=0.5*p*(p+1)  
  pvalue=1-pchisq(test,df) ##p-value of the test statistic  
  crit=qchisq(1-a,df) ##critical value of the chi-square distribution  
  list(test=test,degres=df,p.value=pvalue,critical=crit) }  
combCOV<-matrix(nrow=5,ncol=5,0)
```