

Validity of the Beck Depression Inventory-II using Exploratory and Confirmatory Factor
Analysis among Low-Income Women Enrolled in a Welfare Transition Program

Shawn M. Kneipp, PhD, ARNP
Associate Professor
University of Florida
College of Nursing
Department of Healthcare Environments & Systems

John A. Kairalla, PhD
Assistant Professor
University of Florida
College of Medicine
Department of Epidemiology and Health Policy Research
Division of Biostatistics

Jeanne Stacciarini, PhD
Assistant Professor
University of Florida
College of Nursing
Department of Healthcare Environments & Systems

Deidre Pereira, PhD
Assistant Professor
University of Florida
College of Public Health & Health Professions
Department of Clinical and Health Psychology

Acknowledgements

The authors gratefully acknowledge women in the Welfare Transition Program who participated in this study, and the National Institutes of Health/National Institute of Nursing Research, contract grant number R01NR009406, which supported this research. They also thank Pam Selby for her editorial assistance and comments.

MeSH Keywords: depression, reliability and validity, disadvantaged populations

Abstract

1
2 Background: The Beck Depression Inventory – II (BDI-II) is considered a gold standard for
3 identifying depression in adults. Validity of the BDI-II has been documented in diverse
4 populations using EFA, although no findings have been reported among exclusively lower-
5 income women. Among EFA findings, the BDI-II’s factor structure has been inconsistent, with
6 cognitive, affective, and somatic domains emerging differentially within factors across studies.
7 This, in conjunction with concerns around the confounding of depressive symptoms as measured
8 by the BDI-II and other illness states, have lead researchers to examine more complex factor
9 structures using confirmatory factor analysis (CFA). Objective: We sought to evaluate the factor
10 structure of the BDI-II, using both EFA and CFA among low-income women. Method: EFA with
11 Promax rotation, followed by CFA testing of several structural models with two randomly split
12 subsamples of 108 and 200 women going through a Welfare Transition Program was conducted.
13 Results: EFA indicated a two-factor structure, with the cognitive and affective domains
14 represented in Factor 1, and somatic items comprising Factor 2. A general factor model, with
15 general Depression and residual Cognitive and Somatic factors, best fit to our data based on
16 several indices (RMSEA=.05, SRMR=.05, WRMR=.69, CFI=.98, TLI=.99) and model
17 difference tests of significance (4 comparisons: $\chi^2 > 24.9$ for all, $p < .001$ for all).
18 Discussion: These findings demonstrate BDI-II measurement is best represented by a complex
19 factor structure among low-income women, and is consistent with findings in other populations.
20 Additional consideration for how a general model factor structure provides potentially new
21 directions for depression measurement may advance science in several areas.

22

Background

1
2 Depression is a leading cause of disability worldwide and is a major public health concern
3 in the United States (U.S. Department of Health and Human Services, 1999). Prevalence
4 estimates for Major Depressive Disorder (MDD) among the general U.S. adult population are
5 16.2% for lifetime and 6.6% in the previous 12 months (Kessler et al., 2003). There are,
6 however, significant sex, race/ethnicity, and socioeconomic status (SES) disparities in the
7 prevalence and incidence of depression. Women are nearly twice as likely to suffer from MDD
8 than men, while racial and ethnic minorities are at approximately 1.5 times higher risk for
9 depression than their non-minority counterparts. In addition, there is a four-fold increase in risk
10 of depression for persons living below the federal poverty level (Kessler et al., 2003). Women
11 receiving public assistance (popularly referred to as “welfare”) fall into the highest risk category,
12 with studies suggesting 22% meet the 12-month MDD criteria (Lennon, Blome, & English,
13 2001).

14 The public health impact associated with untreated depression is far-reaching.
15 Absenteeism and lost productivity associated with depression are estimated to cost the U.S. \$24
16 billion each year (McIntyre & O’Donovan, 2004). Furthermore, depression has emerged as an
17 independent risk factor for the initiation and progression of several chronic diseases – including
18 cardiovascular disease and diabetes mellitus (Lustman & Clouse, 2005; Ramasubbu & Patten,
19 2003; Van der Kooy et al., 2007) – and may contribute to disparities in incidence, prevalence,
20 and mortality within these conditions. Given the scope and significance of depression, the use of
21 appropriate assessment measures is crucial for better screening and detection, monitoring
22 treatment outcomes, and continued scientific pursuits to better understand and ameliorate this
23 disorder.

1 The Beck Depression Inventory – Second Edition (BDI- II) is a widely used measure of
2 depressive symptoms and is considered a valid and reliable instrument for use depression
3 screening in the general population (Beck et al., 1996; Vanheule, Desmet, Groenvynck, Rosseel
4 & Fontaine, 2008). The BDI-II yields a single score with standardized ranges indicating
5 no/minimal depressive symptoms to mild, moderate or severe depressive symptoms. Items in the
6 BDI-II reflect cognitive (C), affective (A), and somatic (S) components of depression (Beck,
7 1996 #12), and map closely onto *Diagnostic and Statistical Manual of Mental Disorders- Fourth*
8 *Edition (DSM-IV)* criteria for a Major Depressive Episode (American Psychiatric Association,
9 2000). In methodological studies, these three components are consistently represented in factor
10 analysis findings. What has been substantially less consistent is the alignment of specific BDI-II
11 items onto these components. Some researchers have reported a two-factor structure, with C and
12 A symptoms loading onto one factor and S symptoms loading onto a second factor (“CA-S”
13 structure) (Beck, Steer, & Brown, 1996) while others have reported S and A items loading onto
14 one factor and C items loading onto a second factor (“SA-C” structure) (Arnau, Meagher, Norris,
15 & Bramson, 2001; Beck et al., 1996; Gary & Yarandi, 2004; Steer, Ball, Ranieri, & Beck, 1999).
16 Research has also supported a three-factor structure (“C-A-S”) (Buckley, Parker, & Heggie,
17 2001). Not only have factor structure findings differed, but items representing the factors and
18 the weight of item loadings within factors have also varied across studies. Taken together, these
19 findings suggest that the BDI-II may be comprised of a more complex factor structure than
20 previously reported (Ward, 2006).

21 Several recent methodological studies have explored the potential for more a more
22 complex factor structure of the BDI-II using confirmatory factor analysis (CFA). Findings from
23 these studies generally indicate that second-order (Buckley et al., 2001; Grothe et al., 2005) and

1 general-factor model structures (Thombs, Ziegelstein, Beck, & Pilote, 2008; Ward, 2006) are
2 better fits for the BDI-II than simple two- or three-factor structures. These more complex factor
3 structures may be especially good at modeling BDI-II data among medical populations, for
4 whom the somatic symptoms of depression may be confounded with health status or treatment
5 effects. Few studies, however, have examined the factor structure of the BDI-II in low SES
6 groups (Buckley et al., 2001; Gary & Yarandi, 2004; Grothe et al., 2005). This is a significant
7 gap in the literature, given that low SES individuals are disproportionately affected by
8 depression. Among these studies, only two conducted a CFA, and each study yielded a unique
9 factor structure. Buckley and colleagues (2001) reported a three-factor structure, while Grothe
10 and colleagues (2005) reported a second-order factor structure. To our knowledge, no studies
11 have confirmed the factor structure of the BDI-II in a sample comprised exclusively of low-
12 income women using CFA.

13 Therefore, the purpose of this study was to examine the factor structure of the BDI-II in a
14 sample of low-income women with chronic health conditions using a two-step EFA and CFA
15 approach to validate the underlying construct dimensions for this population, and to contribute to
16 the measurement literature necessary to advance health disparities research.

17 Method

18 *Sample*

19 Participants in this study included 308 low-income women enrolled in an ongoing
20 randomized controlled trial (RCT) to address health disparities among women receiving public
21 assistance benefits through the Temporary Assistance for Needy Families (TANF) program. The
22 study was approved by the university Health Science Center Institutional Review Board.
23 Women met eligibility criteria for enrollment in the ongoing study if they were: receiving TANF

1 benefits, between 18 and 60 years of age, not currently employed, not receiving disability
2 benefits, not pregnant, could read and speak English, and had at least one chronic health
3 condition (either by self-report or identified by a registered nurse during an extensive physical
4 and mental health assessment). A battery of standardized, valid and reliable health measures
5 were administered during the baseline screening health assessment process, including measures
6 of depressive symptoms using the Beck Depression Inventory-II (BDI-II). Data were collected
7 between February 27, 2006 and September 24, 2008. The original sample ($n = 308$) was divided
8 into two subsamples of $n = 108$ to conduct the EFA, and $n = 200$ to conduct the CFA using a
9 computer generated random number sequence utility in Stata/SE Version 10.0. There were no
10 differences between the two randomly generated samples on a number of demographic, health
11 status, or other health-related measures (see Table 1). Mean BDI-II scores were 17.1 and 18.5 in
12 the EFA and CFA subsamples, respectively ($p = 0.37$), indicating mild depressive
13 symptomatology (Beck et al., 1996).

14 *Sample Size Adequacy*

15 A minimum of 5:1, and preferably 10:1, subject-to-item ratio has been a traditional
16 standard for exploratory factor analysis sample size requirements (EFA) (Gorsuch, 1983). More
17 recently, empirical findings using Monte Carlo simulations have advanced sample size
18 estimation methodology for conducting EFA. Specifically, MacCallum and colleagues (2001;
19 1999) demonstrated that, where communalities are moderate to high (≥ 0.40), when there are a
20 small number of factors, and when factors are overdetermined (6 or more indicators per factor), a
21 sample size of 100 (and in some cases less) is sufficient, regardless of the subject-to-item ratio.
22 As with sample size estimation for EFA, traditional sample size recommendations for
23 confirmatory factor analysis (CFA) have ranged between 100 to 200 subjects; however, recent

1 findings suggest sample size adequacy is dependent on key qualities of the CFA model, such as
2 the number of indicators per factor, item loadings, and the number of factors included (Gagne &
3 Hancock, 2006). A goal of sample size estimation for CFA is achieving satisfactory model
4 convergence with the fewest iterations (<1,100 iterations are considered acceptable) (Gagne &
5 Hancock, 2006). With respect to our analyses, the EFA subsample ($n = 108$) met the 5:1 subject-
6 to-item ratio sample size criterion, had moderate-to-high communalities ranging from .40 to .70
7 on >70% of the indicators, and was adequately determined, indicating it met more recent EFA
8 sample size criteria as well. Likewise, the CFA subsample ($n = 200$) met the conventional
9 standard of 200 subjects, and met model convergence with <1,100 iterations, given the indicator
10 per factor structure of the model based on the guidelines of Gagne & Hancock (2006).

11 *Missing data*

12 Of the $n = 308$ women enrolled, 89% ($n = 273$) had complete BDI-II data. The remaining
13 11% ($n = 35$) had missing data at the item-level. Among this relatively small proportion with
14 missing data, the vast majority (88%) had < 2 missing items out of the 21 items comprising the
15 BDI-II. There were no differences in participants with respect to age, race, marital status,
16 number of children, number of chronic conditions, or BDI-II score by complete vs. incomplete
17 case data status, suggesting data were missing at random (Little & Rubin, 2002). Missing data
18 were imputed at the item level using a multinomial logistic regression prediction model (Horton
19 & Laird, 2001), with predictor selection based on the distribution of depressive symptoms by
20 select demographics and bivariate correlations of the BDI-II with other health-related measures.

21 *Exploratory Factor Analysis (EFA) of the BDI-II*

22 *EFA Analysis.* Exploratory factor analysis was performed using Stata Intercooled Version
23 10.0 (StataCorp, 2008). Findings from the developers of the BDI-II (Beck et al., 1996) and the

1 majority of findings conducted by others (Arnau et al., 2001; Gary & Yarandi, 2004; Steer,
2 Rissmiller, & Beck, 2000) indicate the BDI-II is comprised of two symptom dimensions. The
3 goal of this EFA was to examine factor structure in a specific population as opposed to applying
4 it as a data reduction technique for instrument development purposes. Given this, and the
5 expectation that items were not normally distributed, principal factor analysis (PFA) was
6 selected over other common EFA methods such as principal components analysis or maximum
7 likelihood methods (Brown, 2006). Based on prior findings, factor extraction was constrained to
8 a two-factor solution. Results of Bartlett's test of sphericity ($\chi^2 = 1226, p < 0.001$) and the Kaiser-
9 Meyer-Olkin measure of sampling adequacy (0.91) indicated the input matrix was suitable for
10 conducting factor analysis (Pett, Lackey, & Sullivan, 2003). Prior BDI-II EFA findings suggest
11 the factor structure is most clearly interpreted when factors were found to be intercorrelated at
12 approximately 0.70 following oblique rotation (Beck et al., 1996; Thombs et al., 2008; Ward,
13 2006). Drawing upon these findings, a Promax oblique rotation was applied with kappa (κ) set
14 at 4 to best replicate the factor intercorrelation of 0.70 reported by others, allow for maximum
15 separation of factor loadings, and improve interpretability (Brown, 2006). Promax rotation
16 based on $\kappa=4$ resulted in an actual factor correlation of 0.76 and produced the simplest factor
17 solution compared to rotations based on a κ of 3 or 5.

18 *EFA Results.* Factor selection was guided by a set of standard criteria, including: 1) the
19 Kaiser-Guttman rule, 2) the scree plot, 3) cumulative and unique percent of explained variance,
20 4) prior EFA findings, and 5) parallel analysis findings. Based on these criteria, a two-factor
21 solution was most appropriate for our sample data. Eigenvalues were 9.40 and 0.78 for the first
22 and second factor, respectively. If the Kaiser-Guttman rule were the only criterion applied, a
23 one-factor solution would be supported, given only the first factor has an eigenvalue of >1.0 .

1 However, this criterion is to be used with caution when principal components analysis (PCA) is
2 not the extraction method applied, as there is a risk of over- or underestimating the correct
3 number of factors (Pett et al., 2003). The scree plot suggested a two-factor solution. Factor one
4 (F1) explained 81% of the variance extracted, with factor two (F2) explaining an additional 7%.
5 While no definitive guidelines exist for factor selection based on cumulative variance explained,
6 typically a factor contributing $\geq 5\%$ variance should be considered for retention if warranted for
7 theoretical reasons, as is the case for the BDI-II (Pett et al., 2003). Finally, findings from parallel
8 analysis – a less common but useful eigenvalue-based procedure to guide factor selection
9 (Brown, 2006) – supported a two-factor solution as well (Figure 1). With parallel analysis,
10 factor selection is based on the number of eigenvalues from the sample data greater than those
11 generated by random data, indicating a selected factor accounts for more variance than that
12 accounted for by chance alone (Brown, 2006).

13 Factor pattern loadings are presented in Table 2. Factor 1 (F1) was comprised of
14 symptoms belonging to the Cognitive and Affective domains (CA), with symptoms such as self-
15 dislike (.81), self-criticalness (.78), sadness (.77), loss of interest (.75), and guilty feelings (.70)
16 loading highest on this factor. Factor 2 (F2) represented predominantly somatic symptoms (S),
17 including changes in sleep (.72) changes in appetite (.69), irritability (.65), concentration
18 difficulty (.65), tiredness/fatigue (.53), and loss of interest in sex (.47). All 21 items
19 demonstrated acceptable factor loadings (≥ 0.30) on a given factor following rotation, with the
20 vast majority loading at 0.40 or higher (Pett et al., 2003). One item (indecisiveness) dual-loaded
21 on both factors, with loadings of 0.41 on F2 (Somatic, or S) and 0.36 on F1 (Cognitive/Affective,
22 or CA). Seven additional items had no zero loadings on one of the two factors following
23 rotation, indicating the BDI-II is not comprised of a simple factor structure (Gorsuch, 1983).

1 *Item Analysis*

2 The Cronbach alpha of the full BDI-II (items 1-21) was 0.94. There was a high degree of
3 internal consistency within each of the extracted factors, with alpha coefficients of 0.92 in the
4 CA, and 0.84 in the S domains, respectively. Corrected item-total correlations ranged from 0.48
5 – 0.80 for the total BDI-II, 0.50 – 0.83 for F1, and 0.56 – 0.82 for F2.

6 *Confirmatory Factor Analysis (CFA) of the BDI-II*

7 *CFA Analysis.* Confirmatory factor analysis (CFA) was performed using Mplus Version 5.1
8 software (Muthen & Muthen, (1998-2007)). Model selection for evaluation and fit comparison
9 was based on our EFA findings and those findings published by others (Thombs et al., 2008;
10 Ward, 2006). Our EFA findings support a two-factor solution with Cognitive/Affective (CA)
11 and Somatic (S) domains. As noted previously, other CFA findings of the BDI-II indicate a
12 general factor model that initially loads all items on a General factor (G). The remaining variance
13 on residual, orthogonal factors – Somatic (S) and Cognitive (C) (G-S-C model) – provides a
14 better fit over two-factor, first-order models, which have varied in composition across study
15 samples (CA-S, and alternately, SA-C) (Thombs et al., 2008; Ward, 2006). In each of these CFA
16 studies, correlated errors for select item pairs were permitted and improved overall model fit.
17 These correlated error terms included “loss of energy” with “tiredness/fatigue,” “self-dislike”
18 with “self-criticalness,” “loss of pleasure” with “loss of interest,” and “agitation” with
19 “irritability.” Given that the G-S-C model provided the best fit in other populations, that
20 correlated error terms made theoretical sense, and that correlated error terms improved model fit,
21 we opted to examine model fit for different versions of the CA-S and G-S-C models – all of
22 which were specified *a priori* based on our EFA results. The CA-S and G-S-C model versions

1 included: 1) allowing and prohibiting correlated errors on select item pairs, and 2) differentially
2 placing the dual-loading item (indecisiveness) on F1 or F2 in the CA-S models.

3 Given the non-normal distribution of indicators, a robust weighted least squares estimation
4 approach was applied. The chi-square test statistic to evaluate overall model fit was not used as
5 the sole or predominant test of fit, as it is likely to reject a good fitting model due to trivial
6 differences between the correlations and covariances in the observed and predicted matrices
7 (Meyers, Gamst, & Guarino, 2006). Rather, more recently developed and empirically supported
8 indices that examine three general areas of model fit were used. These include measures of: 1)
9 absolute fit (the standardized root mean square residual, or SRMR, and the weighted root-mean-
10 square residual, or WRMR), 2) fit adjusted for parsimony (the root mean square error of
11 approximation, or RMSEA), and 3) comparative or incremental fit (the comparative fit index, or
12 CFI, and the Tucker-Lewis Index, or TLI) (Brown, 2006). Based on the findings of Monte Carlo
13 simulations under varying model conditions, a combined criteria of $>.96$ for TLI or CFI and an
14 SRMR $<.09$, and a WRMR $<.95$ were applied as detailed by Hu and Bentler (1999) and Yu
15 (2002), respectively. Finally, to compare fit between models, χ^2 difference tests of significance
16 were conducted. Consistent with CFA methodology guidelines in the literature, these
17 comparisons were planned in the event more than one model demonstrated acceptable fit based
18 on the TLI/CFI, SRMR, and WRMR indices, and were *not* conducted on revised models
19 informed by our CFA results.

20 *CFA Results.* Model fit statistics are detailed in Table 3. All models demonstrated
21 acceptable goodness of fit based on descriptive fit indices. As described by Brown (2006),
22 however, descriptive fit indices applied in CFA are intended to reflect a model's *lack of fit* as
23 opposed to a model's *degree of fit*. That is, descriptive fit indices are best interpreted as

1 indicators that a model does or does not have an acceptable fit to the data, and are not measures
2 to be used to determine “best fit” among models that all meet minimum goodness of fit criteria.
3 Given the fit indices suggested correlated error terms improved fit in the CA-S model with
4 “indecisiveness” assigned to the CA factor (CA-S1), the fit of CA-S2 (with “indecisiveness”
5 assigned to the S factor) was conducted only using correlated error terms.

6 Examination of the standardized estimates and modification indices suggested localized
7 areas of strain within and across models. Standardized parameter estimates varied from 0.04 to
8 0.85, with the vast majority of items loading at >0.30 (see Table 4). Parameter estimates for the
9 correlated error between “agitation” and “irritability” were <0.30 and not significant in any of
10 the models ($p>0.05$), indicating the amount of shared variance between these indicators
11 contributes little to overall model fit (Brown, 2006). Similarly, in both G-S-C models, estimates
12 for “irritability” on the S factor were <0.30 and not significant ($p>.05$). There was a somewhat
13 wide range of variance explained in the items by the factors, with R^2 estimates between 0.28 and
14 0.72 across models. Within each of the models, the items “changes in appetite” and “loss of
15 interest in sex” had the least amount of variance accounted for ($< 40\%$) by the respective factors.
16 Modification indices (>3.8) suggested improvements in model fit may be obtained in models
17 CA-S1 Correlated, CA-S2 Correlated, and G-S-C (Uncorrelated) by reassigning correlated error
18 terms and reassigning “irritability” and “loss of energy” to the C and S domains, respectively.
19 Correlated error terms suggested in the modification indices differed from those correlated *a*
20 *priori* based on theoretical considerations and empirical support from the findings of Thombs
21 (2008) and Ward (2006). These indices, however, can all be considered marginal – that is, they
22 minimally exceed the modification index cutoff of 3.8 used in Mplus, and would likely
23 contribute little to improving overall model fit (Brown, 2006). Modification indices did not

1 suggest the G-S-C (Correlated) model could be improved; rather, a number of changes were
2 identified that, if made, would substantially worsen model fit.

3 To compare the fit of solutions between CFA models, chi-square difference tests were
4 conducted, whereby a more constrained model (e.g., one with fewer parameters and no error
5 terms allowed to correlate, or H0) is nested within a less constrained model (e.g., one with more
6 parameters due to error terms allowed to correlate, or H1) to determine if one model fit is
7 statistically better than the other (Brown, 2006; Muthen & Muthen, (1998-2007)). In this case, a
8 statistically significant chi-square would indicate the less constrained model (H1) has
9 significantly better fit than the more constrained model (H0). In general, difference tests
10 demonstrated statistically significant improvement in model fit when approached in a step-wise
11 fashion, with sequential improvement noted when moving from models with uncorrelated to
12 correlated error terms, and from the CA-S1 to G-S-C model ($p < 0.001$ for all, see Table 5).
13 Ultimately, consistent with the overall fit indices, pattern coefficient loadings, and modification
14 indices across models, the G-S-C Correlated model demonstrated a statistically better fit with the
15 data above other models evaluated (Figure 2).

16 Discussion

17 This study examined the factor structure of the BDI-II using a two-step EFA and CFA
18 approach in a sample of low-income women with chronic health conditions transitioning from
19 welfare to work. Findings from the EFA indicated that a two factor solution comprised of CA
20 and S domains represented the underlying constructs of the BDI-II in this sample. However, this
21 factor solution was not as well supported when compared against a general-factor (G-S-C) model
22 using CFA. The G-S-C Correlated model was validated to best fit the data, indicating the factor

1 structure of the BDI-II in this population is more complex than previously thought. As in several
2 other studies, the internal reliability of the BDI-II was high, with a Cronbach alpha of .94.

3 These findings build upon the model specification advances published by others and
4 support a more complex factor structure of the BDI-II (Buckley et al., 2001; Grothe et al., 2005;
5 Thombs et al., 2008; Ward, 2006). As noted previously, Buckley et al. (2001) and Grothe et al.
6 (2005) validated a second-order, or “hierarchical” model of the BDI-II, while Ward (2006), and
7 Thombs et al. (2008) found a general-factor model fit as well or better than first-order only
8 models comprised of the varying SA-C and CA-S factor structures typically generated by EFA.
9 With the use of a general factor model approach, our findings are most comparable to those
10 published by Ward (2006) and Thombs (2008), with all BDI-II items initially loading on a
11 general factor (G), and the remaining variance loading on two orthogonal, residual factors (C and
12 S).

13 There are distinct advantages for using a general factor model approach to validate the
14 BDI-II that have implications for interpreting depression data and advancing state-of-the-science
15 depression measurement. First, for those items with a high affinity for inconsistent loading onto
16 factors across study samples, it is difficult to determine their significance in contributing to the
17 overall construct of depression via either the C, A, or S components. For example, the item
18 “indecisiveness” loads inconsistently on the C, A, or S domains when examined using EFA
19 (Arnau et al., 2001; Buckley et al., 2001; Gary & Yarandi, 2004; Steer et al., 1999; 2000). In
20 this study, there were only marginal differences in the factor pattern coefficient loadings of
21 “indecisiveness” on the CA and S factors (.41 and .37, respectively) even post-rotation, calling
22 into question whether this item substantively represents either domain. Moreover, CFA fit
23 indices of the two CA-S models with “indecisiveness” differentially assigned to the CA or S

1 construct were identical, suggesting this item was not unique to either. As found by others
2 (Thombs et al., 2008; Ward, 2006), in our G-S-C model, “indecisiveness” loaded substantively
3 on the General Depression factor (.74), without being forced to contribute to depression via a C,
4 A, or S latent construct, as would be the case with hierarchical, or second-order models (Brown,
5 2006).

6 Second, a general factor model facilitates developing a depressive symptom score capable
7 of partitioning out residual variance in the C and S domains that may not, in fact, reflect
8 depression. For example, as others have pointed out (Thombs et al., 2008), a modified score for
9 the BDI-II may be appropriate for patients with post acute myocardial infarction (MI), whereby
10 the residual variance reflected on S item scores (as with the item “fatigue”) may better represent
11 symptoms due to an illness other than depression. Recent efforts to disentangle somatic
12 symptoms from general depression scores using the Centers for Epidemiologic Studies-
13 Depression (CES-D) scale have taken this approach. Among older persons, Yang et al. (2008)
14 found select chronic conditions are more highly associated with endorsement of psychosomatic
15 items, and they refined CES-D symptom measures based on a general CFA model to reduce
16 confounding.

17 Third, not only do select BDI-II items load inconsistently across factors – individual items
18 and items representing varying “symptom clusters” emerge differently as the most salient item or
19 items within factors across EFA and CFA findings and study populations. Within factors, an
20 item with higher loadings considered representative of a particular factor (Hair et al., 2005), and
21 there are often clusters of items within factors based on the weight of their loadings (Gorsuch,
22 1983). The cognitive domain of the BDI-II, for example, is comprised of items reflecting
23 negative views of self (i.e., self-dislike, self-criticalness, worthlessness), as well as more global

1 items representing pessimism and guilt, among others. Items loading most highly on these
2 factors vary across samples in both EFA and CFA studies. Among EFA results, in a study with
3 rural, older African American women, “pessimism” loaded most highly on the cognitive factor
4 (.81) (Gary & Yarandi, 2004), and “past failure” was most salient for two samples of psychiatric
5 outpatients (.81, .71) (Beck et al., 1996; Steer et al., 1999). In this sample of low-income women,
6 “self-dislike” dominated the cognitive domain (.81), followed closely by “self-criticalness” (.78).
7 Among CFA general factor model results, a symptom cluster consistent with negative views of
8 self shared more variance with General Depression, and much less variance with the C residual
9 factor in our study sample than that reported in MI patients (Thombs et al., 2008) and college
10 student samples (Ward, 2006). Certainly, differences in methods can account for fluctuations in
11 salient loadings and symptom clusters across groups (Gorsuch, 1983). It is also worth
12 considering, however, that depressive symptoms manifest differently across groups by way of
13 salient item clusters, and that they could be conceptually meaningful for understanding the role
14 of depression on relevant health outcomes.

15 In summary, a G-S-C model best represents the factor structure of the BDI-II among low-
16 income women in this sample, and is consistent with findings in other populations.
17 Conceptualizing the structure of the BDI-II using a general factor model approach could
18 contribute to refining the measurement and scoring of depression for use in select populations.
19 In turn, a revised scoring scheme may be particularly applicable for studies examining the affect
20 of depression on morbidity and mortality in the context of specific disease states, and to assess
21 how different manifestations of depression may be important for understanding the
22 psychophysiological mechanisms believed to underlie disease development and health
23 disparities. Given this, continued methodological work in the area of depression measurement is

- 1 needed to allow for more precise hypothesis testing and advancing scientific progress related to
- 2 depression, health outcomes, and health disparities research.

Table 1

Comparison of EFA and CFA Randomly Generated Samples, on Select Characteristics

Characteristic [‡]	EFA Sample (n=108)			CFA Sample (n=200)			<i>p</i> [*]
	%	Mean	SD	%	Mean	SD	
Age		28.9	7.3		29.9	8.0	0.28
Race							
Black	55%			56%			0.76
White	42%			41%			
Other	4%			3%			
Ethnicity							
Hispanic	5%			4%			0.78
Education Level							
<12 th Grade	34%			36%			0.97
High School Diploma/GED	26%			26%			
Some College/College Degree	40%			38%			
College Degree							
Number of Children		2.6	1.4		2.5	1.3	0.49
Mean Age of Children							
Marital Status							
Unmarried	91%			91%			0.95
Married	9%			9%			
Total Household Income		\$520	\$377		\$576	\$451	0.27
Number of Chronic Conditions		2.7	1.7		2.5	1.6	0.47
Total Reported Symptoms		7.8	6.5		8.3	6.9	0.58
PSS [0-56]		30.1	7.8		29.8	6.9	0.70
ISEL [0-120]		52.6	8.2		54.2	7.9	0.14
CFS [1-19]		14.1	4.4		14.8	4.4	0.17
PHQ-9 [0-27]		10.5	6.9		10.7	6.4	0.80
BDI-II [0-63]		17.1	12.0		18.5	11.8	0.37
SF-12 GH [0-100]		41.5	21.5		40.6	23.2	0.74

*Independent t-tests and chi-square applied as appropriate to interval/categorical data, using Fisher's exact test for matrices with <5 observations per cell.

[‡] [] = Possible Score Range; PSS=Perceived Stress Scale; ISEL = Interpersonal Support Evaluation List; CFS = Chronic Financial Strain; PHQ-9 = Patient Health Questionnaire – 9 Item, Symptom Severity Score; BDI-II = Beck Depression Inventory – II; SF-12 GH = Short Form – 12 General Health.

Table 2

Rotated Factor Pattern Matrix Using Promax Rotation, Kappa (κ) = 4

Item Number	Item Description	Factor		h^2
		1	2	
1	Sadness	0.770	-0.045	0.454
2	Pessimism	0.644	0.112	0.468
3	Past Failure	0.583	0.242	0.398
4	Loss of Pleasure	0.663	0.078	0.479
5	Guilty Feelings	0.700	-0.079	0.583
6	Punishment Feelings	0.688	0.020	0.505
7	Self-Dislike	0.810	0.006	0.336
8	Self-Criticalness	0.780	-0.067	0.462
9	Suicidal Thoughts	0.484	0.014	0.755
10	Crying	0.668	-0.013	0.565
11	Agitation	0.269	0.331	0.689
12	Loss of Interest	0.750	0.050	0.379
13	Indecisiveness	0.410	0.365	0.482
14	Worthlessness	0.669	0.176	0.351
15	Loss of Energy	0.458	0.252	0.561
16	Changes in Sleeping	0.001	0.721	0.478
17	Irritability	0.214	0.652	0.327
18	Changes in Appetite	-0.210	0.690	0.687
19	Concentration Difficulty	0.044	0.646	0.539
20	Tiredness/Fatigue	0.213	0.528	0.512
21	Loss of Interest in Sex	0.043	0.466	0.751

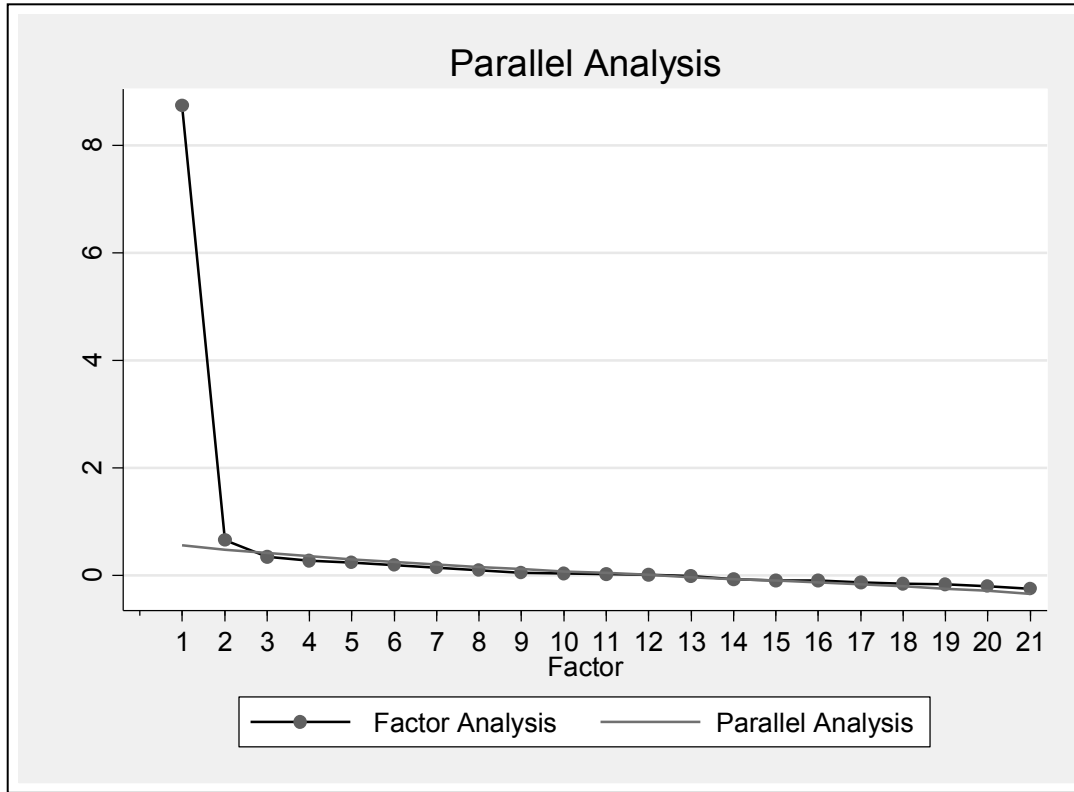


Figure 1. Parallel analysis of BDI-II EFA*

*Based on 10 replications using Stata 10/SE. Eigenvalue of second factor from EFA falls above parallel analysis reference line, suggesting variance accounted for is more than that expected by chance alone.

Table 3

Summary of BDI-II Confirmatory Factor Analysis Results

Model	χ^2	<i>df</i>	<i>P</i>	CFI	TLI	RMSEA	SRMR	WRMR
CA-S1 Uncorr	143.9	67	<0.001	0.949	0.984	0.076	0.060	0.985
CA-S1 Corr	123.6	66	<0.001	0.962	0.988	0.066	0.057	0.797
CA-S2 Corr	123.8	66	<0.001	0.962	0.988	0.066	0.057	0.797
G-S-C Uncorr	114.6	65	0.001	0.967	0.989	0.062	0.051	0.738
G-S-C Corr	100.3	64	<0.05	0.976	0.992	0.053	0.049	0.685

CA-S1 Uncorr: No correlated error terms; “indecisiveness” (Y13) assigned to the Cognitive/Affective domain.

CA-S1 Corr: Cognitive/Affective-Somatic model with correlated error terms on items 4 and 12 (loss of pleasure, loss of interest); 7 and 8 (self-dislike, self-criticalness); 11 and 17 (agitation, irritability); and 15 and 20 (loss of energy, tiredness/fatigue); “indecisiveness” assigned to the Cognitive/Affective domain.

CA-S2 Corr: Cognitive/Affective-Somatic model with same correlated error terms used in CA-S1 Corr; “indecisiveness” assigned to the Somatic domain.

G-S-C Uncorr: General-Somatic-Cognitive model with no correlated error terms allowed.

G-S-C Corr: General-Somatic-Cognitive model with correlated error terms used in CA-S1 and CA-S2.

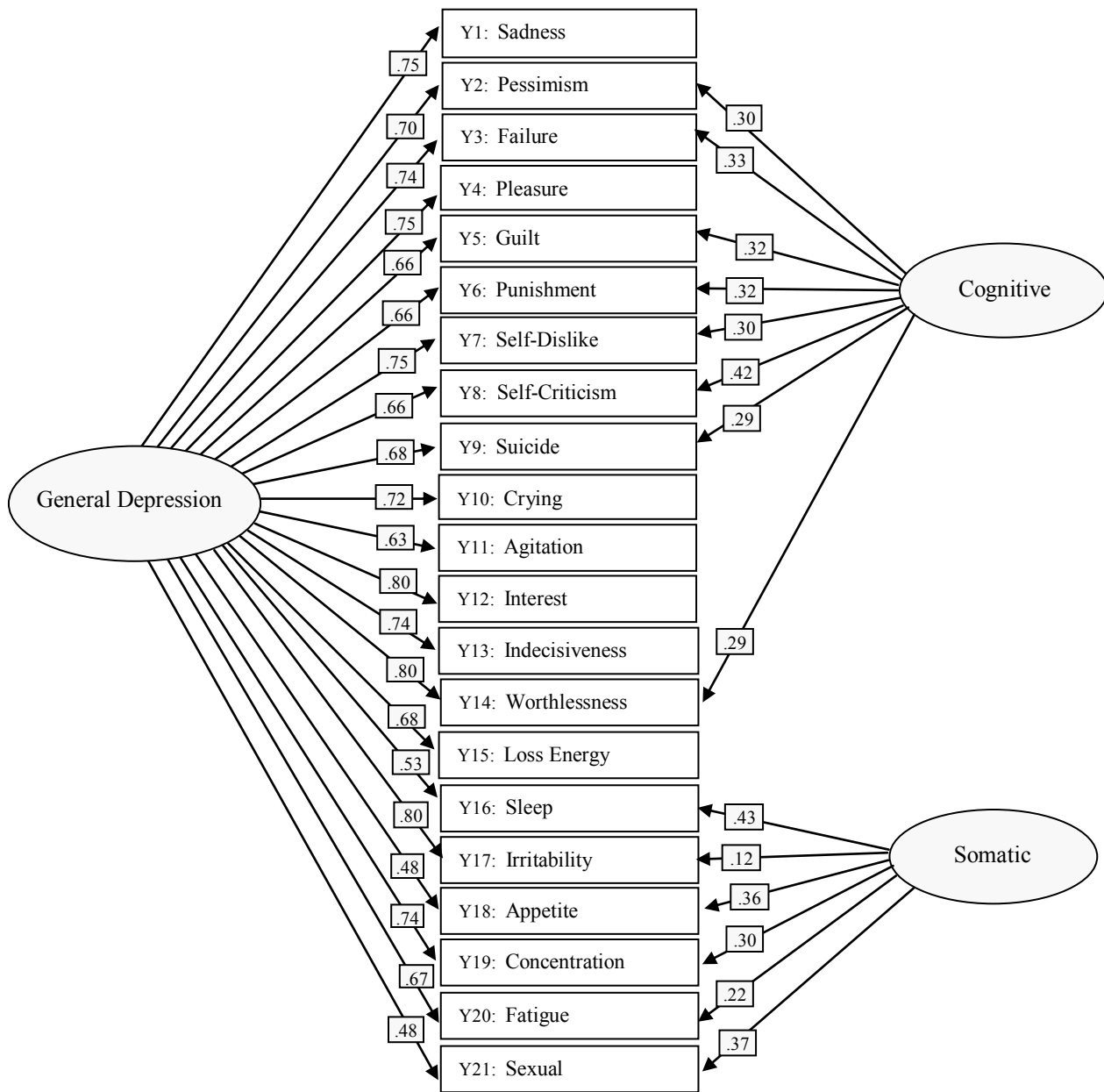


Figure 2. Pattern coefficients for General-Somatic-Cognitive (Corr) Model

Table 4

CFA Standardized Pattern Coefficients and Modification Indices

Model	Loading Range	Loadings <.30	$p > 0.05$ Loadings (Actual p-value)	R^2 Range (Items <.30)	Modification Recommendations
CA-S1 Uncorr	0.53 – 0.92	None	None	0.28 – 0.70 (0.28 = Y21) (0.28 = Y18)	+ S by Y15 - S by Y8
CA-S1 Corr	0.04 - 0.84	0.04 Y11+Y17 0.19 Y4+Y12 0.29 Y7+Y8	Y11 with Y17 ($p=0.65$)	0.28 - 0.71 (0.28=Y18) (0.28=Y21)	+ Y8 with Y6 + Y2 with Y14 + Y4 with Y15 + Y16 with Y18 + Y13 with Y19 - Y8 with Y21
CA-S2 Corr	0.07 - 0.85	0.08 Y11+Y17 0.18 Y4+Y12 0.29 Y7+Y8	Y11 with Y17 ($p=0.36$)	0.28 - 0.71 (0.28=Y18) (0.28=Y21)	+ Y8 with Y6 + Y2 with Y14 + Y16 with Y18 - Y8 with Y21
G-S-C Uncorr	0.10 - 0.81	0.10 Y17 by S 0.17 Y20 by S 0.28 Y2 by C 0.29 Y9 by C 0.29 Y14 by C	Y17 by S ($p=0.28$)	0.36 - 0.72	+ C by Y17 - C by Y21 + S by Y15 - S by C - C by S - S with C
G-S-C Corr	0.12 - 0.80	0.12 Y17 by S 0.14 Y11+Y17 0.19 Y7+Y8 0.16 Y4+Y12	Y17 by S ($p=0.18$) Y11 with Y17 ($p=0.07$)	0.36 - 0.72	- C by Y21 - C on S - S by C - S on C - C by S - S with C

CA-S1 Uncorr: No correlated error terms; “indecisiveness” (Y13) assigned to the Cognitive/Affective domain.

CA-S1 Corr: Cognitive/Affective-Somatic model with correlated error terms on items 4 and 12 (loss of pleasure, loss of interest); 7 and 8 (self-dislike, self-criticalness); 11 and 17 (agitation, irritability); and 15 and 20 (loss of energy, tiredness/fatigue); “indecisiveness” assigned to the Cognitive/Affective domain.

CA-S2 Corr: Cognitive/Affective-Somatic model with same correlated error terms used in CA-S1; “indecisiveness” assigned to the Somatic domain.

G-S-C Uncorr: General-Somatic-Cognitive model with no correlated error terms allowed.

G-S-C Corr: General-Somatic-Cognitive model with correlated error terms used in CA-S1 Corr and CA-S2 Corr.

Table 5

Between-Model Chi-Square Difference Testing Results

Models Tested (H1, H0)*	χ^2	<i>df</i>	<i>p</i> -value
CA-S1 Corr (H1)	45.7	4	<0.001
CA-S1 Uncorr (H0)			
G-S-C Uncorr (H1)	53.3	10	<0.001
CA-S1 Uncorr (H0)			
G-S-C Corr (H1)	43.8	10	<0.001
CA-S1 Corr (H0)			
G-S-C Corr (H1)	24.9	3	<0.001
G-S-C Uncorr (H0)			

*H1 = more restrictive model; H0 = less restrictive model. H0 is

nested within H1 for Mplus chi-square difference testing

(DIFFTEST). Significant *p*-value implies H1 provides better model

fit over H0.

References

- American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th - Text Revision ed.). Washington, DC: American Psychiatric Association.
- 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 Psychol*, *20*(2), 112-119.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II Manual* (2nd ed.). San Antonio: Harcourt Brace & Company.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.
- Buckley, T. C., Parker, J. D., & Heggie, J. (2001). A psychometric evaluation of the BDI-II in treatment-seeking substance abusers. *Journal of Substance Abuse Treatment*, *20*(3), 197-204.
- Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, *41*(1), 65-83.
- Gary, F. A., & Yarandi, H. N. (2004). Depression among southern rural African American women: a factor analysis of the Beck Depression Inventory-II. *Nurs Res*, *53*(4), 251-259.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- 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.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2005). *Multivariate Data Analysis*. 6th Edition. Upper Saddle River, NJ: Pearson Prentice Hall.
- Horton, N. J., & Laird, N. M. (2001). Maximum likelihood analysis of logistic regression models with incomplete covariate data and auxiliary information. *Biometrics*, *57*(1), 34-42.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55.
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., et al. (2003). The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*, *289*(23), 3095-3105.
- Lennon, M. C., Blome, J., & English, K. (2001). *Depression and Low-Income Women: Challenges for TANF and Welfare-to-Work Policies and Programs*. New York, NY: Research Forum on Children, Families and the New Federalism, National Center for Children in Poverty, Mailman School of Public Health, Columbia University.
- Little, R. J., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). Hoboken, New Jersey: Wiley-Interscience.
- Lustman, P. J., & Clouse, R. E. (2005). Depression in diabetic patients: The relationship between mood and glycemic control. *Journal of Diabetes Complications*, *19*(2), 113-122.
- MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*(4), 611-637.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Measurement*, *4*(1), 84-89.
- McIntyre, R. S., & O'Donovan, C. (2004). The human cost of not achieving full remission in depression. *Canadian Journal of Psychiatry*, *49*(Suppl. 1), 10S - 16S.

- Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied Multivariate Research: Design and Interpretation*. Thousand Oaks: SAGE Publications.
- Muthen, L. K., & Muthen, B. O. (1998-2007). *Mplus User's Guide* (5th ed.). Los Angeles: Muthen & Muthen.
- Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. Thousand Oaks: SAGE Publications.
- Ramasubbu, R., & Patten, S. B. (2003). Effect of depression on stroke morbidity and mortality. *Canadian Journal of Psychiatry, 48*, 250-257.
- StataCorp, L. (2008). Stata Intercooled Version 10.0. College Station, TX.
- 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.
- Steer, R. A., Rissmiller, D. J., & Beck, A. T. (2000). Use of the Beck Depression Inventory-II with depressed geriatric inpatients. *Behaviour Research and Therapy, 38*(3), 311-318.
- 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*(2), 115-121.
- U.S. Department of Health and Human Services. (1999). *Mental Health: A Report of the Surgeon General*. Rockville, MD: U.S. Department of Health and Human Services, substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health.
- Van der Kooy, K., van Hout, H., Marwijk, H., Marten, H., Stehouwer, C., & Beekman, A. (2007). Depression and the risk for cardiovascular diseases: Systematic review and meta analysis. *International Journal of Geriatric Psychiatry, 22*(7), 613-626.
- Ward, L. C. (2006). Comparison of factor structure models for the Beck Depression Inventory--II. *Psychological Assessment, 18*(1), 81-88.
- Yang, F. M., & Jones, R. N. (2008). Measurement differences in depression: Chronic health-related and sociodemographic effects in older Americans. *Psychosomatic Medicine, 70*(9), 993-1004.
- Yu, C. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcome*. Unpublished Dissertation, University of California. Los Angeles, Los Angeles.